AD_____

Award Number: DAMD17-00-1-0287

TITLE: Hormone Receptors in Breast Cancer Prognosis - Racial and Quantitative Effects

PRINCIPAL INVESTIGATOR: Carl M. Tammemagi, Ph.D.

CONTRACTING ORGANIZATION: Henry Ford Health System
Detroit, Michigan 48202

REPORT DATE: June 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2003	Annual (1 Jun 02-31 May 03)

4. TITLE AND SUBTITLE
Hormone Receptors in Breast Cancer Prognosis - Racial and Quantitative Effects

5. FUNDING NUMBERS
DAMD17-00-1-0287

6. AUTHOR(S)
Carl M. Tammemagi, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Henry Ford Health System
Detroit, Michigan 48202

E-Mail: Mtammem1@HFHS.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Breast cancer survivors compose the largest group of cancer survivors in the United States. As heterogeneity exists within stages and between races in breast cancer survival, it is important to develop a better understanding of prognostic factors. Tumor estrogen and progesterone receptors are one of the more important prognostic factors in breast cancer patients. However, currently in clinical practice hormone receptor status is treated as either being present or absent and is treated similarly in all groups. The dichotomization of hormone status may lead to loss of valuable information and hormone receptor status may not have the same effect in African Americans and Whites. This historical cohort study evaluates quantitative differences in tumor hormone receptors in African Americans and Whites and determines whether survival effects differ between the two groups. This study also assesses whether a dose-response relationship, linear or nonlinear, exists between hormone receptors and survival. Findings of this study may lead to better prediction of survival and to identification of subsets of patients at higher risk that may have gone unrecognized by the application of a single cutpoint. Our preliminary findings indicate that African American breast cancer patients have more estrogen receptor negativity and a worse survival.

14. SUBJECT TERMS
Breast cancer survival, African American, estrogen receptor, progesterone receptor, quantitative analysis, comorbidity, socioeconomic status

15. NUMBER OF PAGES
56

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

Breast cancer survivors compose the largest group of cancer survivors in the United States today. As considerable heterogeneity exists within stages and between racial groups in breast cancer survival, it is important to develop a better understanding of prognostic factors. Estrogen and progesterone receptors in breast tumor tissue are regarded to be one of the more important prognostic factors in breast cancer patients. However, currently in clinical practice hormone receptor status is treated as either being present or absent and is treated similarly in all race/ethnic groups. The dichotomization of hormone status may lead to loss of valuable information and hormone receptor status may not have the same effect in African Americans and whites. This historical cohort study evaluates quantitative differences in estrogen and progesterone receptors in the breast tumors of African Americans and whites and determines whether survival effects differ between the two groups. This study will also assess whether a dose-response relationship, linear or nonlinear, exists between quantitatively assessed hormone receptors and survival, as opposed to the currently popular dichotomized assessment of receptor status. Findings of this study may lead to better prediction of survival and to identification of subsets of patients needing particular clinical attention that may have gone unrecognized by applying one cutpoint to all patients.

BODY

The majority of study tasks has been accomplished and is described in the *Statement of Work* described in Table 1.

Table 1. Progress on items in the *Statement of Works*

	Description	*Planned time*	*Progress*
Task 1	Initial establishment of study team, approach and issues	1 to 4 months	Completed
	Staff training	1 to 4 months	Completed
	Preparation of computer programs and study database	1 to 4 months	Completed
Task 2	Establish and Characterize Cohort	4 to 8 months	Completed
	Abstraction of Patient / Tumor Data From	4 to 16 months	
	Computer Databases		Completed
	Medical Record Abstraction		Completed
	Hormone Receptor Log Book		Completed
Task 3	SES estimates based on 1990 US Census data	12 to 24 months	In progress
Task 4	Survival Data Collection From		
	Henry Ford Health System Tumor Registry	12 to 24 months	Completed
	SEER	18 to 24 months	Completed
	Michigan State Tumor Registry	18 to 24 months	Completed
Task 5	Attend breast cancer conference	Year 2 and 3	1 Attended, DoD ERA of HOPE 2002 attended
Task 6	Analysis, preparation of manuscripts and reports	Year 3	In progress

The last report detailed the abstraction of medical records and hormone receptor data from clinical logs and the entry of data into Microsoft Access databases. This work is now complete. Data clean-up and analysis is currently in progress. Two obstacles have lead to the study

extending beyond its planned three year duration: (1) inadequate measurement of socioeconomic status (SES), and (2) processing of the large volume of comorbidity data.

Measurement of SES. In our last report we described our strategy for estimating SES using *block group median household income* (BGMHI) derived from the 1990 US census using patients' addresses. We were able to obtain BGMHI SES estimates for only 70 percent of study subjects. This is unsatisfactory. In addition, in one of our recent studies we demonstrated that other area-based socioeconomic measures were more strongly correlated with individual education, which is regarded as one of the best single individual estimators of SES. We found that *proportion employed in managerial or professional occupations* was the single most powerful predictor of individual education at block group, tract and zip code levels. Others have recommended using tract or block group *proportion under poverty level* as the single most appropriate area-based socioeconomic measures [1, 2]. Additionally, we found in the Detroit population that some of the area-based socioeconomic measures interacted significantly with gender or race. These findings have been submitted for publication (Appendix 1) and bear directly on the analysis of the current study. To ensure maximum possible assessment of our study population and to make possible evaluation of a large meaningful set of area-based SES estimators, we recently acquired *MapInfo Professional* ® v7.0 and *MapMarker* ® v.8.1 software programs (MapInfo Corporation, Troy, NY), in conjunction with Spatial Re-Engineering Consultant's (SRC) *Portfolio Desktop* ® (Orange, CA), the latter being a sophisticated data retrieval engine for demographic statistics. This system is expected to give us the power to carry out high quality and advanced SES analyses. While this software system came initially loaded to allow immediate evaluations using the 2000 US census data, the 1990 US census data, which we

had purchased, had inadvertently not been included in the system. This problem is being rectified. Our computer programmer, Richard Krajenta, is currently taking a training course on this system and is obtaining technical support in the application of this system to 1990 US census data. Within a month we expect to be able to evaluate multiple SES estimators in the vast majority of our study population.

Comorbidity. We have recently completed a comprehensive study of the impact of comorbidities in lung cancer survival [3-5], which included a comprehensive evaluation of factors accounting for race/ethnic disparity in survival. In that study, more African American survival disparity was accounted for by adverse comorbidities (23.2%) and adverse symptoms (46.3%) than was by stage (19.2%) (manuscript in preparation). This important, novel finding was made possible by detailed abstraction of medical records, which is not possible in exclusively Tumor Registry-based studies, which include the majority of past studies, and also by the relatively high power delivered by the elevated proportion of blacks in our population. These favorable conditions are also present in the current study. We have taken recently discovered knowledge from our lung cancer study and have integrated it into our Department of Defense study to extend it beyond that originally proposed:

1. The impact of comorbidity on breast cancer survival will be evaluated in a much more comprehensive fashion than has previously been carried out, for example by the popular, but simplistic, Charlson Index [6].

2. The impact of adverse comorbidities and symptoms in explaining African American race/ethnic differences in treatment and survival will be evaluated, and if found to be important the association between hormone receptor status and survival will be adjusted for them.

3. The association between comorbidities, in particular obesity, diabetes, lipid problems and thyroid/glandular diseases and hormone receptor status will be evaluated. These evaluations make *a priori* sense because race/ethnic differences in the distribution of these comorbidities as well as hormone receptor status exist and the associations between these comorbidities and carcinogenesis have been postulated [7-10]

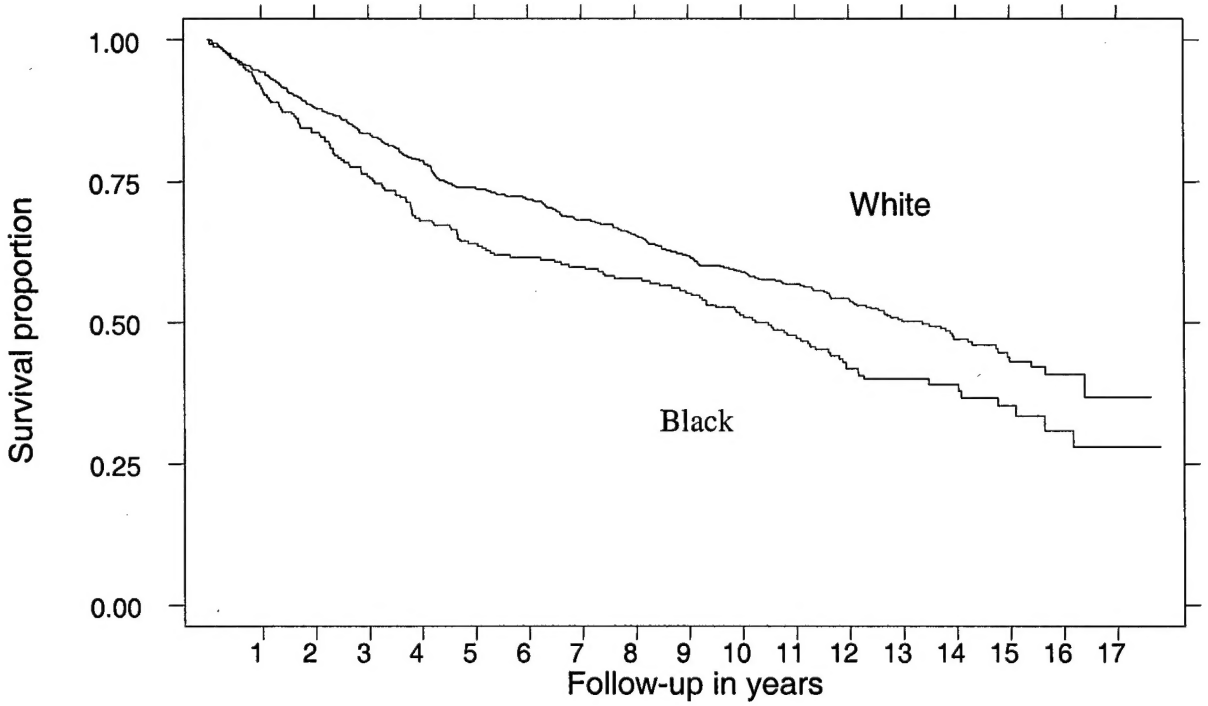
This extension of data collection and analysis (>350 additional variables, Appendix 2) in part explains the delay in completion of the study. However, we expect the results to reflect the additional effort.

Dr. Tammemagi presented the following poster at the Era of Hope meeting in September 2002: **Tammemagi CM**, Neslund-Dudas C, Feldkamp C. *Hormone receptors and breast cancer prognosis – Racial and quantitative effects.* Era of Hope (Department of Defense Breast Cancer Research Meeting) September 25-28, 2002, Orlando, Florida.

Preliminary Study Findings

The median follow-up of the breast cancer cohort was 12.1 years. The 5-year survival for African Americans was 0.64 (95% CI 0.57, 0.69) and for whites was 0.74 (95% CI 0.70, 0.77). The hazard ratio (black vs. white) was 1.34 (95% CI 1.10, 1.63; p = 0.004). A Kaplan Meier survival plot describing the survival experience for these two groups is presented in Figure 1.

Figure 1. Kaplan Meier survival plot describing the survival experience of breast cancer patients, HFHS, diagnosed 1985-1990, by race/ethnicity.



KEY RESEARCH ACCOMPLISHMENTS: NA.

REPORTABLE OUTCOMES: NA.

CONCLUSIONS

Although the study has been extended beyond the originally proposed time period, we expect it to be enriched with detailed analyses of SES and comorbidities, well beyond that originally planned. We expect all analyses to be complete by June 2004 and manuscripts and final report to be submitted prior to December 2004.

REFERENCES

1. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health* 2003;93(10):1655-71. Available from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14534218.

2. Singh GK, Miller AB, Hankey BF, Edwards BK, Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975-1999. NIH Publication No. 03-5417. National Cancer Institute, 2003.

3. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Impact of comorbidity on lung cancer survival. *Int J Cancer* 2003;103(6):792-802. Available from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12516101.

4. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Smoking and lung cancer survival - the role of comorbidity and treatment. *Chest* 2004;125(1):27-37.

5. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. In lung cancer patients, age, gender, race-ethnicity and smoking predict adverse comorbidity, which in turn predicts lung cancer treatment & survival. *Journal of Clinical Epidemiology* 2004;(in press).

6. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373-83.

7. Kaaks R. Nutrition, hormones, and breast cancer: is insulin the missing link? *Cancer Causes Control* 1996;7(6):605-25. Available from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8932921.

8. Manjer J, Kaaks R, Riboli E, Berglund G. Risk of breast cancer in relation to anthropometry, blood pressure, blood lipids and glucose metabolism: a prospective study within the Malmo Preventive Project. *Eur J Cancer Prev* 2001;10(1):33-42. Available from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11263589.

9. Kaaks R, Lukanova A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. *Proc Nutr Soc* 2001;60(1):91-106. Available from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11310428.

10. Lukanova A, Lundin E, Toniolo P, Micheli A, Akhmedkhanov A, Rinaldi S, et al. Circulating levels of insulin-like growth factor-I and risk of ovarian cancer. *Int J Cancer* 2002;101(6):549-54. Available from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12237896.

APPENDIX 1

Associations between Individual and Aggregate Measures of Socioeconomic Status in Detroit – Interactions with Gender & Race

Dr. C. Martin Tammemagi (Author for correspondence and reprint requests)
Senior Epidemiologist
Josephine Ford Cancer Center
Henry Ford Health System
1 Ford Place, 5C
Detroit, Michigan 48202-3450
phone: 313 874-4640
fax: 313 874-6656
email: mtammem1@hfhs.org

Dr. Lee Richard Kallenbach
Assistant Professor
Wayne State University
Department of Community Medicine
Detroit, Michigan
phone: 313 577-6679
fax: 313 577-0316
email: lkallenb@med.wayne.edu

Abbreviations:
CI, confidence interval;
SD, standard deviation
SES, socioeconomic status

Key words: aggregate area-based socioeconomic measures, socioeconomic status, gender, race/ethnicity, interaction

Running heading: Individual and Aggregate Measures of SES

ACKNOWLEDGEMENT

The authors are grateful to the following individuals who generously provided data for this study: Drs. Christine Cole-Johnson, Marvella Ford, Jay Gorell, Robert Morlock, David Nerenz, and Benjamin Rybicki, Patricia Williams, and Ms. Suzanne Havstad and Christine Neslund-Dudas. Data analyzed in this study was acquired under the support of grants and funding from the Schering-Plough Pharmaceutical Company, the Fund for Henry Ford Hospital, the National Institute of Environmental Health Sciences (ES 06418) and the National Institutes of Allergy and Infectious Diseases (AI24156).

ABSTRACT

Area-based socioeconomic measures (ABSM) are often employed in place of individual socioeconomic status (SES) data or in combination with the latter in multilevel analyses. This study evaluates the relationship between ABSM and individual SES at different area levels and interactions by gender and race/ethnicity. Bootstrap correlation coefficients and linear regression analysis were used to evaluate associations between individual education and 21 ABSMs at census block group, census tract and zip code levels for 1789 subjects participating in five Metropolitan Detroit epidemiologic studies. Associations were strongest at census block group and census tract levels. The correlations between *individual education* and *proportion employed in managerial / professional occupations* (PEMPO) at block group and tract levels were significantly stronger than any other parings (for both, correlation coefficient = 0.42, 95% CI 0.38, 0.46). In multivariate models, predictors of *individual education* were *PEMPO* interacting with *gender*, *date of birth*, and *median household income* (MHI) interacting with *race/ethnicity*. At all three area levels, for PEMPO above the median, women generally had lower education than men and the difference grew with increasing PEMPO. At all three area levels, for any MHI above $20,000, blacks had higher education than their white counterparts and the difference increased with increasing MHI. These interactions were significant in all three area levels. These findings suggest that analyses at the block group and tract level are preferred, that PEMPO should be further investigated as a useful measure of SES, and gender- and race-ABSM interactions need to be considered, especially when explaining disparities using ABSM.

Abstract word count: 252.

Word count including text, references, tables and figures: 6759.

Socioeconomic status (SES) is a complex, multidimensional phenomenon that is a robust, profound predictor of health and disease. Understanding the association between the SES gradient and health requires understanding of its measurement at multiple levels and in different formats.

Area-based socioeconomic measures (ABSM) often employed in the absence of individual SES data or in combination with individual SES data in multilevel analyses. Although our understanding of ABSM is at a nascent stage, evidence is accumulating that supports its use in health research.

Studies by Geronimus and colleagues (Geronimus, Bound, & Neidert, 1996; Geronimus & Bound, 1998) and Soobader and colleagues (Soobader, LeClere, Hadden, & Maury, 2001; Soobader & LeClere, 1999) indicate that aggregate estimators of SES should not serve as "proxies" for their corresponding individual SES measure because biases may be introduced. For example, census area per capita income should not be used to represent an individual's personal income per se. Nevertheless, these and other studies (Balfour & Kaplan, 2002; Duncan, Jones, & Moon, 1993, 1999) do suggest that aggregate SES estimators do stand on their own as important measures of SES. Social advantage and disadvantage are not randomly spatially distributed (Hyndman, Holman, Hockey, Donovan, Corti, & Rivera, 1995) and if the area of analysis is relatively homogeneous and a gradient exists between different areas, then aggregate estimators of SES are expected to be informative. Soobader and colleagues found that aggregate SES estimators explained as much variation in self-perceived health as did regression models using individual SES (R^2 = 21% for both models) (Soobader, LeClere, Hadden et al., 2001). Many racial differences in health have in large part been attributable to differences in SES. In regression models, Soobader and colleagues found that aggregate SES estimators consistently

"explained away" more of the race-health association than did individual SES estimators (Soobader, LeClere, Hadden et al., 2001; Soobader & LeClere, 1999). Anderson et al. using National Longitudinal Mortality Study data found that median census tract household income was significantly associated with increased risk of 11-year mortality in shites and blacks of both genders, after adjustment for individual-level income (Anderson, Sorlie, Backlund, Johnson, & Kaplan, 1997). ABSM may capture a broader picture of an individual's SES because ABSM correlate with numerous individual measures of SES and in addition may incorporate community or contextually relevant features of SES (Berkman & Kawachi, 2000; Diez Roux, 2002; Geronimus, Bound, & Neidert, 1996; Geronimus & Bound, 1998; Krieger, 1992; Soobader, LeClere, Hadden et al., 2001; Soobader & LeClere, 1999).

ABSM do not suffer from some weaknesses inherent in particular individual level estimators. Requests for income data can offend and often go missing in surveys. This is problematic at extremes where important differences in effects may exist. Income does not always reflect the acquired wealth of an individual or family. Occupation as a measure of SES, although useful (Moss & Krieger, 1995), can be difficult to stratify. The social strata of specific occupations can fluctuate in different cohorts and periods, and the classification of homemakers and retirees can be troublesome. Education is considered by some to be the best single measure of SES (Berkman & Macintyre, 1997), because education is available for most study subjects, is stable over time and is unaffected by sickness and temporary unemployment and is reflective of the SES of homemakers and retirees. Soobader et al. found that education was the single best individual level predictor of self-perceived health (Soobader, LeClere, Hadden et al., 2001).

ABSM can provide valuable information for health researchers. Lack of standard, valid measures of SES in vital statistics, government and non-government health surveys, and disease

registries prompted the National Institutes of Health conference *Measuring Social Inequalities in Health* (Annapolis, MD, September, 1994) to recommend the utilization of census-based aggregate SES estimators in the absence of individual SES data (Moss & Krieger, 1995). However, appropriate application and interpretation of aggregate estimators of SES have not been established. At this time it is uncertain as to what size of aggregate area is optimal for estimating SES, which aggregate SES variables are most informative, and whether such estimators serve equally well in different age, gender and race groups (Liberatos, Link, & Kelsey, 1988) and geographic regions. The construct of SES is not a simple one and the relationships between different measures of SES are not expected to be necessarily linear. As a consequence it is important to evaluate interactions and nonlinear effects (Adler, Boyce, Chesney, Cohen, Folkman, Kahn et al., 1994).

To address some of these issues, we carried out a study in Metropolitan Detroit evaluating the relationship between individual SES (education) and a variety of ABSM at the census block group, census tract, and zip code levels. The specific aims were to determine (i) which area level best predicts individual education; (ii) which specific ABSM best predict individual education within aggregate levels; and (iii) whether the relationship between individual education and ABSM differs by gender and race.

METHODS

Data from 2,703 subjects from five epidemiologic studies carried out in the Henry Ford Health System in the mid-1990's were pooled. The Henry Ford Health System is a large vertically integrated health system that in 1996 was responsible for the health care of approximately 460,000 individuals, 16 percent of Metropolitan Detroit's population. The

subjects for this study were cases and controls in a study of Parkinson's disease (Gorell, Johnson, Rybicki, Peterson, & Richardson, 1998), mothers of children enrolled in an asthma study (Joseph, Ownby, Peterson, & Johnson, 2000), and subjects involved in studies of asthma, diabetes (Nerenz, Repasky, Whitehouse, & Kahkonen, 1992) and back pain (Morlock, Nerenz, Benzel, Nockels, Dempsey, Enwood et al., 2002). For the majority of subjects, age, race, gender, education and individual address data, collected through personal interview, were available. All studies received Institutional Review Board approval. Excluded from the pooled study were individuals who were not African American or White, who did not live in Metropolitan Detroit (Wayne, Macomb and Oakland counties), who were under 25 years of age, and individuals for whom complete education and address data were unavailable. Individuals were entered into the analysis only once.

Individual educational attainment was grouped into seven ordinal levels: (i) grade eight or less, (ii) some high school but grade 12 not completed, (iii) high school graduate, (iv) some college or associate degree less than a Bachelor of Science or Arts degree equivalent, (v) college or university bachelor degree equivalent completed, (vi) some graduate or professional school, but not completed, and (vii) post-graduate or professional degree attained. From individuals' addresses at time of study and 1990 US census data, 25 aggregate variables (Table 1) were derived at block group, tract and zip code levels. Aggregate and individual data were collected blind to each other.

Statistical Methods

The correlations between individual and ABSM were evaluated using Pearson's correlation coefficients (r). Correlation coefficients and 95% confidence intervals (CI) were

estimated using bootstrap methods with 1000 re-samplings per estimate (Chernick, 1999; Efron & Tibshirani, 1993). Whether two correlation coefficients were significantly different was tested using the Fisher R-to-Z transformation test (Neal, 2000). Linear regression modeling was used to determine which ABSM predict individual education within each area levels, adjusted for relevant sociodemographic variables. Non-linear associations were evaluated by multivariate adaptive regression spline (MARS) analysis (Friedman & Roosen, 1995; Zhang & Singer, 1999). SAS 6.12 (SAS Institute Inc., Cary, NC), S-plus 6 (Insightful Inc., Seattle, WA) and Stata 7 (Stata Corporation, College Station, TX) software were used to prepare statistics, models and figures.

RESULTS

The number of analytic subjects was 1789: 554 male (31.0%), 1235 female (69.0%), 261 Black (14.6%) and 1528 Whites (85.4%). The mean age of study participants was 48.3 years (SD 21.8). The mean individual education level was 3.90 (SD 1.34). The distribution of individual education is described in Figure 1. The mean individual education for women was 3.90 (SD 1.23) and for men was 3.91 (SD 1.58) (t-test $p = 0.84$), and for Blacks was 3.63 (SD 1.54) and for Whites was 3.95 (SD 1.30) (t-test $p = 0.002$). The distribution of aggregate variables at the three area levels is presented in Table 1.

Correlations between individual education and aggregate SES variables

Table 2 presents correlation coefficients and 95% confidence intervals between individual education and aggregate SES estimators at the three area levels. Table 3 presents the more important of these correlations further stratified by gender and race/ethnicity. In all levels,

the strongest correlation between *individual education* and ABSM was with *proportion employed in managerial/professional occupations* (PEMPO), and the strength of this association was significantly stronger in the block group and tract levels (both $r = 0.42$, 95% CI 0.38, 0.46) than in the zip code level ($r = 0.37$, 95% CI 0.33, 0.41) (Fisher R-to-Z test $p = 0.01$ comparing block group or tract to zip code level). At the block group level, the mean *PEMPO* for individuals in each of the seven ascending education categories was 0.21, 0.24, 0.27, 0.31, 0.38, 0.41 and 0.45.

The second strongest correlations in the block group and tract levels were between *individual education* and *proportion of individuals ≥16 years with a high school diploma*: r was 0.35 (95% CI 0.31, 0.39) at the block group level and 0.37 (95% CI 0.33, 0.41) at the tract level. These correlations were significantly smaller than the correlations between *individual education* and *PEMPO* (block group level comparison $p = 0.002$, tract level comparison $p < 0.001$). At the zip code level, $r_{\text{individual education~\% high school diploma}}$ was 0.32 (95% CI 0.27, 0.36) and this correlation was significantly smaller than $r_{\text{individual education~PEMPO}}$ (0.37, 95% CI 0.33, 0.41) ($p = 0.001$).

Median household income (MHI) and *median family income* had similar correlations with *individual education* at all three aggregate levels (r range 0.31-0.33, Table 2) and these correlations were significantly smaller than observed between *individual education* and *PEMPO* (for all comparisons $p < 0.001$).

The relatively strong correlation between *individual education* and *PEMPO* was consistently observed in both race groups but was greater in males than in females (Table 3): at the block group level, r was 0.41 for whites, 0.41 for blacks, 0.38 for women, and 0.48 for men. In contrast, the correlation between *individual education* and *median household income* was

similar in men and women but was higher in blacks than in whites: at the tract level, *r* was 0.28 in whites, 0.43 in blacks, 0.30 in women and 0.35 in men.

Predictors of individual education – multivariate linear regression analysis

Linear regression modeling was carried out with individual education as the dependent variable and sociodemographic and ABSM variables in one aggregate level as the available predictor variables. Model results are presented in Table 4. In all three aggregate levels, *PEMPO* interacting with *gender*, *date of birth*, and *median household income* interacting with *race*, were significant predictors of individual education. The two interactions will be detailed at the block group level.

The *PEMPO*gender* interaction predicting *individual education* at the block group level is presented graphically in Figure 2. Women living in block groups with lower PEMPO on average had higher education levels than men, whereas women living in block groups with higher PEMPO generally had lower education levels than their male counterparts. The crossover point at which men and women had similar education levels occurred where PEMPO equaled 32 percent (54^{th} percentile). The regression beta for 10 percent change in PEMPO in women was 0.33 (95% CI 0.29, 0.38) and in men was 0.49 (95% CI 0.42, 0.57). This gender effect was consistent in unadjusted and adjusted models in all three area levels (Table 5) and in blacks and whites (Figure 2).

The *median household income*race/ethnic* interaction predicting *individual education* association at the block group level is presented graphically in Figure 3. In block groups with mean *MHI* > $20,000, African Americans generally had higher education than their White counterparts, and the education difference increased with as mean MHI became larger. This

effect is evident in both genders (Figure 3). The association between *individual education* and *MHI* was markedly reduced by adjustment for covariates in multivariate models, whereas the association between *individual education* and *PEMPO* was not (Table 5).

MARS modeling demonstrated that *individual education* increased with birth cohorts from the beginning of the 1900's until 1951 and then decreased with subsequent years (Figure 4). In multivariate linear regression analysis, the beta coefficient for *date of birth* until 1951 was 0.25 (95% CI 0.20, 0.29) per decade and from 1951 on was – 0.65 (95% CI – 0.98, – 0.32). Age was not significant in models containing *date of birth.*

DISCUSSION

The strongest correlations between individual and aggregate SES, occurred between *individual education* and *proportion in managerial/professional occupations* at the block group and tract levels (both $r = 0.42$, 95% CI 0.38, 0.46) and these two correlations were significantly stronger than any other correlations evaluated. The next strongest correlations were between *individual education* and *proportion of individuals over 16 years with a high school diploma* at the block group and tract levels ($r = 0.35$ & 0.37). *Median household income* and *median family income* had correlations with *individual education* that ranged between 0.31 and 0.33 in the three aggregate levels.

Multivariate linear regression and MARS (data not shown) analyses confirmed the important independent association between *individual education* and *proportion in managerial/professional occupations.* However, this association was modified significantly by gender and this interaction was observed consistently at all three aggregate levels. Both analytic approaches demonstrated that *date of birth* was an independent predictor of *individual education.*

In addition, in linear regression analysis, *median household income* interacting with *race/ethnicity* significantly predicted *individual education* at all three aggregate levels, and to an important extent this effect was independent of the effect associated with PEMPO.

Regarding which ABSM should be used in research, Geronimus and Bound found that median family income consistently predicted individual SES better than other aggregate measures and they recommend median income as "a sensible single aggregate measure to use" (Geronimus, Bound, & Neidert, 1996; Geronimus & Bound, 1998). Others have recommended or used *proportion below poverty* (Krieger, Chen, Waterman, Rehkopf, & Subramanian, 2003; Singh, Miller, Hankey, & Edwards, 2003). The findings of the current investigation suggest that PEMPO in addition to MHI might be informative. *PEMPO* was significantly stronger at predicting *individual education* than was *median household income* and *median family income.* The predictive power of *PEMPO* remained strong in both genders in all three aggregate levels after adjustment for other relevant predictors. In contrast, the predictive ability of *median household income* was considerably reduced following model adjustment (Table 5) and this decline was almost completely explained by adjustment for *PEMPO* (data not shown).

Regarding which area level should is optimum for research, it has been suggested that "block group data can identify pockets of poverty or affluence not apparent at the tract level" (Krieger, 1992). In the current study, results from correlation, linear regression and MARS analyses found that individual SES~ABSM associations were stronger at the block group and tract levels than at the zip code level. However, there was no sharp distinction between the former two levels, and selected aggregate variables even at the zip code level had moderately

strong correlations. Soobader et al. similarly found that there was only slight difference between block group and tract estimates, although they only assessed two ABSM. Kreiger and colleagues reporting on their Public Health Disparities Project endorse the use of tract level ABSM (Krieger, Chen, Waterman et al., 2003).

The findings of this study indicate that education levels demonstrated a nonlinear cohort effect – it increased over time for those born in the first half of the 20th century and thereafter declined. Whether the observed birth cohort effect is specific to this population is unknown and further evaluation of such effects in this and other populations is needed.

The current study detected interactions for important ABSM by gender and race/ethnicity. The strongest association in this study, between *individual education* and *PEMPO*, differed significantly by gender, with the linear regression beta coefficient being greater in men than in women. One possible explanation for why women in the higher SES communities on average had lower education than their male counterparts is that they interrupted their education because of marriage and/or child rearing. Figure 3 demonstrates that all of those individuals that are at education level 6, incomplete graduate or professional school, were women. To determine if older women catch up educational goals postponed at an earlier age, analysis was restricted to those older than 40 years. This stratification did not change the gender interaction.

The dissimilarity index is a measure of residential segregation and represents the proportion of blacks that are needed to move across census tracts to get a uniform distribution of Black residents across an entire Metropolitan Statistical Area. A dissimilarity index of greater than 0.6 reflects hypersegregation of a city. Detroit, with a dissimilarity index of 0.87 in 1990, is

one of the most segregated cities in the United States (Glaeser & Vigdor, 2001). Because blacks and whites tend to live in different neighborhoods, an important issue is whether individual-aggregate SES associations differ by race. Stratified analysis and modeling of interaction terms indicated that the single strongest predictor of individual education, *PEMPO,* did so comparably in both groups. In contrast, the association between *median household income* and *individual education* differed significantly by race.

The positive associations observed in this study between individual education and *PEMPO* and between individual education and *MHI* are expected, but reasons for their interactions with gender and race/ethnicity are not clear. One possible explanation is that the data are differentially misclassified. Address data for contact purposes are generally well maintained in the health system as well as in study subjects and consequential misclassification of address data is unlikely. Education data is obtained by self-report and has a greater potential for misclassification. The most likely scenario is that those with low educational attainment exaggerate their achievement. The difference in the *individual education~PEMPO* association by gender (Figure 2) is difficult to explain by misreporting. To obtain the observed interaction, at low PEMPO one gender would have to exaggerate or under-report their educational attainment while at high PEMPO that gender would do the opposite.

To obtain the *individual education~MHI* interaction with *race/ethnicity* (Figure 3), blacks would have to have exaggerated their educational attainment and do so increasingly as the MHI rises. Such a pervasive deception is hard to accept, as is the contrary explanation that Whites systematically under-report their education. Selection (participation) bias may explain some of the race/ethnic interaction observed in Figure 3, if black individual's decision to participate was

influenced to a greater extent by their education than it was for Whites. But why this would occur when ABSM is estimated by *PEMPO* and not *MHI* is not explained.

The interactions observed in this study may reflect real associations. It is possible that in communities with low PEMPO the drop-out rate is greater in males while in communities with high PEMPO women have "married up" to be paired with men with higher education or have left the education stream earlier than their male counterparts due to marital, child rearing or other commitments. The interaction between *PEMPO*race/ethnicity* in predicting individual education is compatible with the observation by others that to earn the same income as whites, blacks often require a higher education. National data for 1996 indicate that the median income by educational attainment was lower for Blacks compared to Whites at every education level (Williams, 1999), and data from the current study are consistent with those finding (Figure 5).

Dr. Krieger and colleagues studied the relationship between ABSM and health outcomes stratified by race/ethnicity and gender (Krieger, Chen, Waterman et al., 2003) and concluded that disparities should be monitored "by geocoding US public health surveillance systems and using the census tract-level measure *percentage of persons below poverty*". They state "one advantage of ABSMs is that they can be applied equally to all persons, regardless of age, gender, and employment status ...". The current study suggests that important interactions exist between certain ABSM and race/ethnicity and between ABSM and gender, and thus they should not necessarily be considered to have equivalent effects in different race/ethnic or gender groups. Indeed, examination of Krieger's tabular data (their Table 4), which we presented in our Figure 6, demonstrates that the single ABSM *% below poverty* applied to both blacks and whites would fail to explain race/ethnic differences in absolute number of premature male deaths and apparent differences in slopes describing the associations, i.e., interaction between race/ethnicity and %

below poverty. The findings of the current study suggest that more than one ABSM and interaction terms may improve prediction of outcomes and explanation of race/ethnic disparities by SES. For the data presented in Figure 6, if one evaluated the amount of race/ethnic disparity explained by SES using *% below poverty* as a single variable, would lead to model misspecification and under-estimation of the impact of SES on race/ethnic disparity, because the estimated parameter would be averaged between blacks and whites and would underestimate the higher rate of change observed in blacks.

The current study size is modest in comparison to others that have utilized national data. This may appear to be a disadvantage. However, the current study had adequate power to demonstrate important relationships, including interactions, and excessively large "over-powered" studies have the tendency to find even trivial associations to be significant. Limiting the study to one site has the advantage of increasing homogeneity. Some effects may be regional and pooling disparate regions together may dilute and obscure important associations.

The current study did not sample the general population, but drew its subjects from five heterogeneous medical/epidemiologic studies. Although study participants generally tend to have higher SES than nonparticipants (Giuliano, Mokuau, Hughes, Tortolero-Luna, Risendal, Ho et al., 2000; Harlan, Sandler, Lee, Lam, & Mark, 1995; Rimer, Schildkraut, Lerman, Lin, & Audrain, 1996; Trauth, Musa, Siminoff, Jewell, & Ricci, 2000), we are unaware of compelling reasons indicating that individual-ABSM associations are different in research subjects compared to the general population. At minimal, it is expected that the associations observed in the current study reflect those observed in other research samples, which often rely on aggregate SES data. However, it is reassuring to observe that several statistics presented here are in remarkably close

agreement with those observed in population-based studies. Soobader et al. applying linear regression to National Health Interview Survey data were able to explain 22.5 percent (R^2) of variation in individual education with age, gender, race and aggregate income and education variables at the block group level and 21.5 percent at the tract level (Soobader, LeClere, Hadden et al., 2001). The current study using sociodemographic and ABSM variables explained 22.8 and 22.7 percent variation at the block group and tract levels, respectively. Soobader et al. found that individual and aggregate SES variables were moderately correlated (r = 0.33-0.44) and were similar in the block group and tract levels. Those correlations are comparable to those presented in our Table 2. Geronimus et al. evaluating the *Panel Study of Income Dynamics* data (1) found the correlation between zip code median household income and individual education was r = 0.31 versus r = 0.32 in the current study.

If the study findings are confirmed, the observations that *PEMPO* and *MHI* predict *individual education* independently of each other and that they interact with different factors indicate that these two ABSM measure distinct aspects of SES. With regard to utilization of ABSM in research, this study suggests the following tentative recommendations:

(i) Block group or tract data are preferred over zip code ABSM.

(ii) *Proportion employed in managerial/professional occupations* may be the single best area-based SES estimator.

(iii) As some different ABSM appear to be measuring unique aspects of SES, consideration can be given to modeling more than one ABSM at a time.

(iii) In the study of SES, interactions and non-linear effects should be evaluated routinely.

REFERENCES

Adler, N.E., Boyce, T., Chesney, M.A., Cohen, S., Folkman, S., Kahn, R.L., & Syme, S.L. (1994). Socioeconomic status and health. The challenge of the gradient. *Am Psychol*, 49(1), 15-24.

Anderson, R.T., Sorlie, P., Backlund, E., Johnson, N., & Kaplan, G.A. (1997). Mortality effects of community socioeconomic status. *Epidemiology*, 8(1), 42-47.

Balfour, J.L., & Kaplan, G.A. (2002). Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study. *Am J Epidemiol*, 155(6), 507-515.

Berkman, L., & Macintyre, S. (1997). The measurement of social class in health studies: old measures and new formulations. In M. Kogevinas,N. Pearce,M. Susser, & P. Boffetta (Eds.), *Social Inequalities and Cancer* (pp. 51-64). Lyon, France: International Agency for Research on Cancer.

Berkman, L., & Kawachi, I. (2000). A historical framework for social epidemiology. In L. Berkman, & I. Kawachi (Eds.), *Social Epidemiology* (pp. 3-12). New York: Oxford University Press.

Chernick, M. (1999). *Bootstrap Methods, A Practitioner's Guide* New York: John Wiley & Sons, Inc.

Diez Roux, A.V. (2002). Invited commentary: places, people, and health. *Am J Epidemiol*, 155(6), 516-519.

Duncan, C., Jones, K., & Moon, G. (1993). Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med*, 37(6), 725-733.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: are there neighbourhood effects? *Soc Sci Med*, 48(4), 497-505.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap* New York: Chapman & Hall

Friedman, J.H., & Roosen, C.B. (1995). An introduction to multivariate adaptive regression splines. *Stat Methods Med Res*, 4(3), 197-217.

Geronimus, A., Bound, J., & Neidert, L. (1996). On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association*, 91(434), 529-537.

Geronimus, A.T., & Bound, J. (1998). Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. *Am J Epidemiol*, 148(5), 475-486.

Giuliano, A.R., Mokuau, N., Hughes, C., Tortolero-Luna, G., Risendal, B., Ho, R.C.S., Prewitt, T.E., & McCaskill-Stevens, W.J. (2000). Participation of minorities in cancer research: the influence of structural, cultural, and linguistic factors. *Ann Epidemiol*, 10(8 Suppl), S22-34.

Glaeser, E., & Vigdor, J. (2001). Racial Segregation in the 2000 Census: Promising News. Survey Series April 2001 pp. 1-16). Washington, DC: The Brookings Institution, Center on Urban & Metropolitan Policy.

Gorell, J.M., Johnson, C.C., Rybicki, B.A., Peterson, E.L., & Richardson, R.J. (1998). The risk of Parkinson's disease with exposure to pesticides, farming, well water, and rural living. *Neurology*, 50(5), 1346-1350.

Harlan, W.R., 3rd, Sandler, S.A., Lee, K.L., Lam, L.C., & Mark, D.B. (1995). Importance of baseline functional and socioeconomic factors for participation in cardiac rehabilitation. *Am J Cardiol*, 76(1), 36-39.

Hyndman, J.C., Holman, C.D., Hockey, R.L., Donovan, R.J., Corti, B., & Rivera, J. (1995). Misclassification of social disadvantage based on geographical areas: comparison of postcode and collector's district analyses. *Int J Epidemiol*, 24(1), 165-176.

Joseph, C.L., Ownby, D.R., Peterson, E.L., & Johnson, C.C. (2000). Racial differences in physiologic parameters related to asthma among middle-class children. *Chest*, 117(5), 1336-1344.

Krieger, N. (1992). Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health*, 82(5), 703-710.

Krieger, N., Chen, J.T., Waterman, P.D., Rehkopf, D.H., & Subramanian, S.V. (2003). Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health*, 93(10), 1655-1671.

Liberatos, P., Link, B.G., & Kelsey, J.L. (1988). The measurement of social class in epidemiology. *Epidemiol Rev*, 10, 87-121.

Morlock, R., Nerenz, D., Benzel, E., Nockels, R., Dempsey, P., Enwood, S., Feil, E., Kalfas, I., Krauss, W., Pfeifer, B., Rosenblum, M., & Ward, R. (2002). Spine surgeons accurately estimate the probability of favorable one-year postoperative lumbar surgery outcomes (in press). *Disease Management.*

Moss, N., & Krieger, N. (1995). Measuring social inequalities in health. Report on the Conference of the National Institutes of Health. *Public Health Rep*, 110(3), 302-305.

Neal, D. (2000). Bootstrap inferences about measures of correlation (program 55 sg138). *Stata Technical Bulletin*, 10, 149-152.

Nerenz, D.R., Repasky, D.P., Whitehouse, F.W., & Kahkonen, D.M. (1992). Ongoing assessment of health status in patients with diabetes mellitus. *Med Care*, 30(5 Suppl), MS112-124.

Rimer, B.K., Schildkraut, J.M., Lerman, C., Lin, T.H., & Audrain, J. (1996). Participation in a women's breast cancer risk counseling trial. Who participates? Who declines? High Risk Breast Cancer Consortium. *Cancer*, 77(11), 2348-2355.

Singh, G.K., Miller, A.B., Hankey, B.F., & Edwards, B.K. (2003). Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975-1999. NIH Publication No. 03-5417. Bethesda, MD: National Cancer Institute.

Soobader, M., LeClere, F.B., Hadden, W., & Maury, B. (2001). Using aggregate geographic data to proxy individual socioeconomic status: does size matter? *Am J Public Health*, 91(4), 632-636.

Soobader, M.J., & LeClere, F.B. (1999). Aggregation and the measurement of income inequality: effects on morbidity. *Soc Sci Med*, 48(6), 733-744.

Trauth, J.M., Musa, D., Siminoff, L., Jewell, I.K., & Ricci, E. (2000). Public attitudes regarding willingness to participate in medical research studies. *J Health Soc Policy*, 12(2), 23-43.

Williams, D.R. (1999). Race, socioeconomic status, and health. The added effects of racism and discrimination. *Ann N Y Acad Sci*, 896, 173-188.

Zhang, H., & Singer, B. (1999). *Recursive Partitioning in the Health Sciences* New York: Springer

TABLE 1. Area-based aggregate variable mean values (standard deviation, range) by three aggregate levels

Variable	Census block group	Census tract	ZIP code
Number of persons	1490 (1004, 18-5645)	4599 (1811, 707-10815)	30085 (16027, 1090-84704)
Number of families	407 (273, 8-1548)	1256 (498, 142-2955)	8121 (4,224, 329-22238)
Count of white population	1355 (1011, 0-5198)	4150 (2044, 0-10064)	25,309 (14773, 522-80428)
Count of black population	94 (269, 0-2445)	334 (951, 0-5676)	4029 (11392, 0-65226)
Proportion population ≥ age 25 with high school diploma	0.81 (0.12, 0.20-1)	0.81 (0.11, 0.34-0.99)	0.80 (0.10, 0.44-0.97)
Proportion population ≥ age 25 with high school diploma, whites	0.80 (0.17, 0-1)	0.81 (0.12, 0-1)	0.80 (0.10, 0.34-0.97)
Proportion population ≥ age 25 with high school diploma, blacks	0.33 (0.44, 0-1)	0.56 (0.45, 0-1)	0.78 (0.25, 0-1)
Proportion population ≥ age 16 in labor force that is employed	0.94 (0.06, 0.52-1)	0.93 (0.05, 0.56-0.99)	0.93 (0.05, 0.59-0.98)
Proportion population ≥ age 16 in labor force that is employed, whites	0.92 (0.16, 0-1)	0.93 (0.10, 0-1)	0.94 (0.03, 0.61-0.98)
Proportion population ≥ age 16 in labor force that is employed, blacks	0.34 (0.45, 0-1)	0.57 (0.46, 0-1)	0.87 (0.24, 0-1)
Proportion employed in managerial/professional occupations	0.33 (0.14, 0-0.80)	0.32 (0.13, 0.08-0.68)	0.31 (0.11, 0.12-0.62)
Median household income	$45,016 (18,84, 0-98757)	$43,738 (16823, 0-91991)	$42,466 (14559, 6399-89977)
Proportion households with income <$25,000, whites	0.25 (0.20, 0-1)	0.27 (0.18, 0-1)	0.28 (0.14, 0.06-0.84)
Proportion households with income <$25,000, blacks	0.11 (0.25, 0-1)	0.20 (0.32, 0-1)	0.33 (0.28, 0-1)
Median family income	$49,056 (18529, 0-98820)	$48,338 (17273, 0-98239)	$47,720 (15489, 0-98745)
Per capita income, whites	$19,538 (9978, 0-94357)	$19,287 (8542, 0-80687)	$18,993 (7003, 5511-52709)
Per capita income, blacks	$7,317 (13052, 0-94518)	$11,984 (13655, 0-86852)	$16,769 (12292, 0-92000)
Proportion persons living below the poverty level	0.07 (0.10, 0-0.64)	0.07 (0.09, 0-0.65)	0.08 (0.09, 0.01-0.56)
Proportion persons living below the poverty level, whites	0.06 (0.11, 0-1)	0.06 (0.10, 0-1)	0.07 (0.07, 0.01-0.62)
Proportion persons living below the poverty level, blacks	0.06 (0.17, 0-1)	0.09 (0.21, 0-1)	0.12 (0.16, 0-1)
Number of housing units	567 (401, 8-3184)	1,763 (689, 476-4898)	11,686 (6157, 449-34508)
Number owned housing units	416 (279, 0-1727)	1,297 (542, 0-2933)	8,221 (4145, 82-19929)
Number rented housing units	124 (222, 0-2177)	383 (391, 3-2856)	2,936 (2562, 45-13,175)
Median rent per month for rental housing units	$516 (274, 0-1001)	$565 (175, 0-1001)	$533 (117, 215-934)
Median value of owned housing units	$39,520 (35812, 0-99500)	$40,281 (35561, 0-99600)	$42,613 (35338, 0-99800)

TABLE 2. Bootstrap estimates of correlation coefficients (95% confidence intervals) between *individual education* and aggregate variables at three area levels

Aggregate level variable	Block Group	Tract	ZIP Code
Number of persons	0.11 (0.05, 0.15)	0.07 (0.03, 0.11)	-0.14 (-0.19, 0.09)
Number of families	0.12 (0.07, 0.16)	0.09 (0.05, 0.13)	-0.12 (-0.17, -0.07)
Count of white population	0.13 (0.08, 0.18)	0.11 (0.06, 0.16)	-0.03 (-0.08, 0.02)
Count of black population	-0.12 (-0.17, -0.07)	-0.12 (-0.17, -0.07)	-0.12 (-0.17, -0.07)
% population ≥25 yrs. with high school diploma	0.35 (0.31, 0.39)	0.37 (0.33, 0.41)	0.32 (0.27, 0.36)
% population ≥25 years with high school diploma, whites	0.27 (0.22, 0.31)	0.32 (0.28, 0.36)	0.33 (0.28, 0.37)
% population ≥25 years with high school diploma, blacks	0.09 (0.04-0.14)	0.12 (0.07, 0.16)	0.09 (0.05, 0.14)
% population ≥16 years in labor force employed	0.23 (0.19-0.27)	0.25 (0.21, 0.29)	0.22 (0.17, 0.27)
% population ≥16 years in labor force employed, whites	0.17 (0.12, 0.21)	0.14 (0.10, 0.18)	0.21 (0.17, 0.25)
% population ≥16 years in labor force employed, blacks	0.07 (0.02-0.11)	0.08 (0.03, 0.13)	0.07 (0.02, 0.11)
% employed in managerial/professional occupations	0.42 (0.38-0.46)	0.42 (0.38-0.46)	0.37 (0.33, 0.41)
Median household income	0.33 (0.28, 0.37)	0.32 (0.27-0.36)	0.32 (0.27, 0.36)
% households with income less than $25,000, whites	-0.23 (-0.27, -0.18)	-0.27 (-0.31, -0.23)	-0.28 (-0.32, -0.24)
% households with income less than $25,000, blacks	-0.14 (-0.19, 0.10)	-0.14 (-0.19, -0.10)	-0.22 (-0.27, -0.18)
Median family income	0.32 (0.27, 0.37)	0.31 (0.27, 0.36)	0.31 (0.27, 0.36)
Per capita income, whites	0.32 (0.27, 0.37)	0.35 (0.31, 0.39)	0.34 (0.30, 0.39)
Per capita income, blacks	0.12 (0.07, 0.17)	0.14 (0.09, 0.19)	0.18 (0.13, 0.23)
% below the poverty line	-0.22 (-0.26, -0.18)	-0.22 (-0.26, -0.18)	-0.20 (-0.25, -0.15)
% below the poverty line, whites	-0.15 (-0.19, -0.11)	-0.18 (-0.23, -0.14)	-0.20 (-0.25, -0.15)
% below the poverty line, blacks	-0.09 (-0.14, -0.04)	-0.12 (-0.17, -0.07)	-0.14 (-0.19, -0.09)
Number of housing units	0.10 (0.05, 0.15)	0.06 (0.01, 0.11)	-0.14 (-0.19, -0.10)
Number owned housing units	0.13 (0.08, 0.17)	0.11 (0.06, 0.15)	-0.08 (-0.13, -0.03)
Number rented housing units	0.002 (-0.06, 0.06)	-0.04 (-0.10, 0.01)	-0.17 (-0.22, -0.12)
Median rent for rental housing units	0.09 (0.03, 0.14)	0.27 (0.22, 0.31)	0.31 (0.26, 0.35)
Median value of owned housing units	-0.15 (-0.19, -0.11)	-0.15 (-0.20, -0.10)	-0.12 (-0.17, -0.08)

TABLE 3. Bootstrap estimates of correlation coefficients (95% confidence intervals) between *individual education* and selected area-based socioeconomic measures at three aggregate levels, stratified by gender and race/ethnicity

Area-based socioeconomic measures	Block Group	Tract	ZIP Code
% population ≥ 25 with high school diploma	0.35 (0.31, 0.39)	0.37 (0.33, 0.41)	0.32 (0.27, 0.36)
Females	0.33 (0.28, 0.38)	0.35 (0.29, 0.40)	0.30 (0.25, 0.36)
Males	0.39 (0.31, 0.45)	0.42 (0.36, 0.49)	0.34 (0.27, 0.41)
Whites	0.32 (0.28, 0.37)	0.36 (0.31, 0.40)	0.30 (0.25, 0.35)
Blacks	0.42 (0.31, 0.52)	0.43 (0.31, 0.53)	0.35 (0.23, 0.47)
% of whites ≥ 25 years with high school diploma	0.27 (0.22, 0.31)	0.32 (0.28, 0.36)	0.33 (0.28, 0.37)
Females	0.26 (0.21, 0.31)	0.28 (0.23, 0.33)	0.31 (0.26, 0.36)
Males	0.29 (0.21, 0.37)	0.38 (0.32, 0.44)	0.35 (0.28, 0.42)
Whites	0.32 (0.27, 0.36)	0.36 (0.31, (0.41)	0.32 (0.27, 0.36)
Blacks	0.25 (0.13, 0.36)	0.28 (0.18, 0.37)	0.31 (0.20, 0.42)
% of blacks ≥ 25 years with high school diploma	0.09 (0.04, 0.14)	0.12 (0.07, 0.16)	0.09 (0.05, 0.14)
Females	0.09 (0.03, 0.15)	0.12 (0.06, 0.17)	0.09 (0.03, 0.14)
Males	0.09 (0.01, 0.17)	0.12 (0.04, 0.20)	0.11 (0.04, 0.19)
Whites	0.11 (0.06, 0.16)	0.12 (0.07, 0.17)	0.06 (0.02, 0.11)
Blacks	0.34 (0.20, 0.47)	0.39 (0.27, 0.49)	0.31 (0.19, 0.42)
% employed in managerial/professional occupations	0.42 (0.38, 0.46)	0.42 (0.38, 0.46)	0.37 (0.33, 0.41)
Females	0.38 (0.33, 0.43)	0.37 (0.31, 0.42)	0.33 (0.28, 0.39)
Males	0.48 (0.41, 0.55)	0.50 (0.43, 0.56)	0.41 (0.34, 0.48)
Whites	0.41 (0.37, 0.46)	0.42 (0.37, 0.46)	0.36 (0.31, 0.41)
Blacks	0.41 (0.30, 0.51)	0.41 (0.27, 0.52)	0.36 (0.24, 0.47)
Median household income	0.33 (0.28, 0.37)	0.32 (0.27, 0.36)	0.32 (0.27, 0.36)
Females	0.33 (0.27, 0.38)	0.30 (0.24, 0.35)	0.31 (0.25, 0.36)
Males	0.34 (0.26, 0.41)	0.35 (0.27, 0.42)	0.33 (0.25, 0.41)
Whites	0.30 (0.25, 0.35)	0.28 (0.23, 0.33)	0.29 (0.24, 0.35)
Blacks	0.41 (0.29, 0.52)	0.43 (0.31, 0.54)	0.38 (0.26, 0.49)
Median family income	0.32 (0.27, 0.37)	0.31 (0.27, 0.36)	0.31 (0.27, 0.36)
Females	0.34 (0.29, 0.39)	0.29 (0.23, 0.34)	0.31 (0.26, 0.37)
Males	0.30 (0.21, 0.38)	0.35 (0.27, 0.43)	0.32 (0.23, 0.40)
Whites	0.29 (0.24, 0.34)	0.28 (0.23, 0.34)	0.29 (0.24, 0.34)
Blacks	0.41 (0.29, 0.51)	0.41 (0.29, 0.51)	0.39 (0.26, 0.50)
Per capita income for whites	0.32 (0.27, 0.37)	0.35 (0.31, 0.39)	0.34 (0.30, 0.39)
Females	0.29 (0.23, 0.35)	0.31 (0.25, 0.36)	0.32 (0.26, 0.37)
Males	0.36 (0.27, 0.44)	0.40 (0.33, 0.47)	0.38 (0.30, 0.45)
Whites	0.31 (0.26, 0.37)	0.34 (0.29, 0.39)	0.33 (0.28, 0.38)
Blacks	0.29 (0.17, 0.40)	0.33 (0.21, 0.44)	0.35 (0.23, 0.47)
Per capita income for blacks	0.12 (0.07, 0.17)	0.14 (0.09, 0.19)	0.18 (0.13, 0.23)
Females	0.09 (0.03, 0.15)	0.11 (0.05, 0.17)	0.16 (0.10, 0.23)
Males	0.17 (0.09, 0.24)	0.19 (0.11, 0.27)	0.21 (0.12, 0.29)
Whites	0.12 (0.06, 0.17)	0.12 (0.07, 0.17)	0.15 (0.10, 0.21)
Blacks	0.36 (0.24, 0.47)	0.40 (0.28, 0.50)	0.32 (0.20, 0.43)

TABLE 4. Linear regression models predicting *individual education* at three aggregate levels

Variable in model	Beta (β)	95% CI	P-value	Standardized β *
Block Group Level Model (Adjusted R^2 = 22.8%)				
PEMPO	2.78	2.22, 3.38	<0.001	0.30
Gender	-0.23	-0.53, 0.065	0.12	-0.08
*PEMPO *Gender* Interaction	1.66	0.85, 2.48	<0.001	0.22
Date of birth	45e-6	35e-6, 55e-6	<0.001	0.22
Median household income (MHI)	5.30e-06	1.14e-6, 9.46e-6	0.01	0.075
Race/ethnicity	-0.25	-0.58, 0.069	0.12	-0.067
*MHI * Race/ethnicity* Interaction	12e-06	2.15e-6, 22e-6	0.02	0.097
Intercept	2.96	2.75, 3.16	<0.001	-
Tract Level Model (Adjusted R^2 = 22.7%)				
PEMPO	3.54	2.84, 4.23	<0.001	0.34
Gender	-0.31	-0.63, 0.0071	0.05	-0.11
*PEMPO * Gender* Interaction	1.88	0.99, 2.77	<0.001	0.24
Date of birth	47e-6	37 e-6, 57 e-6	<0.001	0.23
Median household income (MHI)	-1.31e-06	-6.47e-06, 3.84e-06	0.63	-0.017
Race/ethnicity	-0.45	-0.79, -0.11	0.009	-0.12
*MHI * Race/ethnicity* Interaction	18 e-6	7.81e-06, 29 e-6	0.001	0.14
Intercept	3.03	2.81, 3.25	<0.001	-
Zip Code Level Model (Adjusted R^2 = 17.8%)				
PEMPO	3.96	2.96, 4.96	<0.001	0.34
Gender	-0.14	-0.49, 0.22	0.45	-0.047
*PEMPO * Gender* Interaction	1.39	0.36, 2.43	0.008	0.17
Date of birth	47 e-6	36 e-6, 57 e-6	<0.001	0.23
Median household income (MHI)	-4.40e-06	-12 e-6, 3.42e-06	0.27	-0.051
Race/ethnicity	-0.48	-0.85, -0.11	0.01	-0.13
*MHI * Race/ethnicity* Interaction	16 e-6	5.26e-06, 27 e-6	0.004	0.13
Intercept	3.05	2.80, 3.29	<0.001	-

Abbreviations: MHI, median household income; PEMPO, proportion employed in managerial or professional occupations.

* Standardized β describes the amount of change in *individual education* in standard deviations per one standard deviation change in the predictor variable.

TABLE 5. Unadjusted and adjusted linear regression beta coefficients for the predictor variables *proportion in managerial/professional occupations* stratified by gender, and *median household income* stratified by race, at three aggregate levels. Dependent variable: *individual education.*

Stratifying variable: Gender	**Female**		**Male**	
Predictor variable	**Unadjusted β**	**Adjusted β ***	**Unadjusted β**	**Adjusted β ***
PEMPO, Block group	0.33, $p < 0.001$	0.28, $p < 0.001$	0.49, $p < 0.001$	0.43, $p < 0.001$
PEMPO, Tract	0.36, $p < 0.001$	0.34, $p < 0.001$	0.55, $p < 0.001$	0.56, $p < 0.001$
PEMPO, ZIP code	0.38, $p < 0.001$	0.37, $p < 0.001$	0.51, $p < 0.001$	0.57, $p < 0.001$
Stratifying variable: Race/ethnicity	**Whites**		**Blacks**	
Predictor variable	**Unadjusted β**	**Adjusted β †**	**Unadjusted β**	**Adjusted β †**
MHI per $20,000, Block group	0.44, $p < 0.001$	0.09, $p = 0.04$	0.78, $p < 0.001$	0.46, $p = 0.001$
MHI per $20,000, Tract	0.47, $p < 0.001$	-0.06, $p = 0.28$	0.87, $p < 0.001$	0.51, $p = 0.001$
MHI per $20,000, ZIP code	0.55, $p < 0.001$	-0.14, $p = 0.09$	0.79, $p < 0.001$	0.38, $p = 0.04$

Abbreviations: MHI, median household income; PEMPO, proportion employed in managerial or professional occupations.

* Adjusted for *MHI*, *Race/ethnicity* and *MHI * Race/ethnicity* interaction, and *Date of birth.*

† Adjusted for *PEMPO*, *Gender*, and *PEMPO * Gender* interaction, and *Date of birth.*

FIGURE 1. The distribution of educational attainment in the population under study

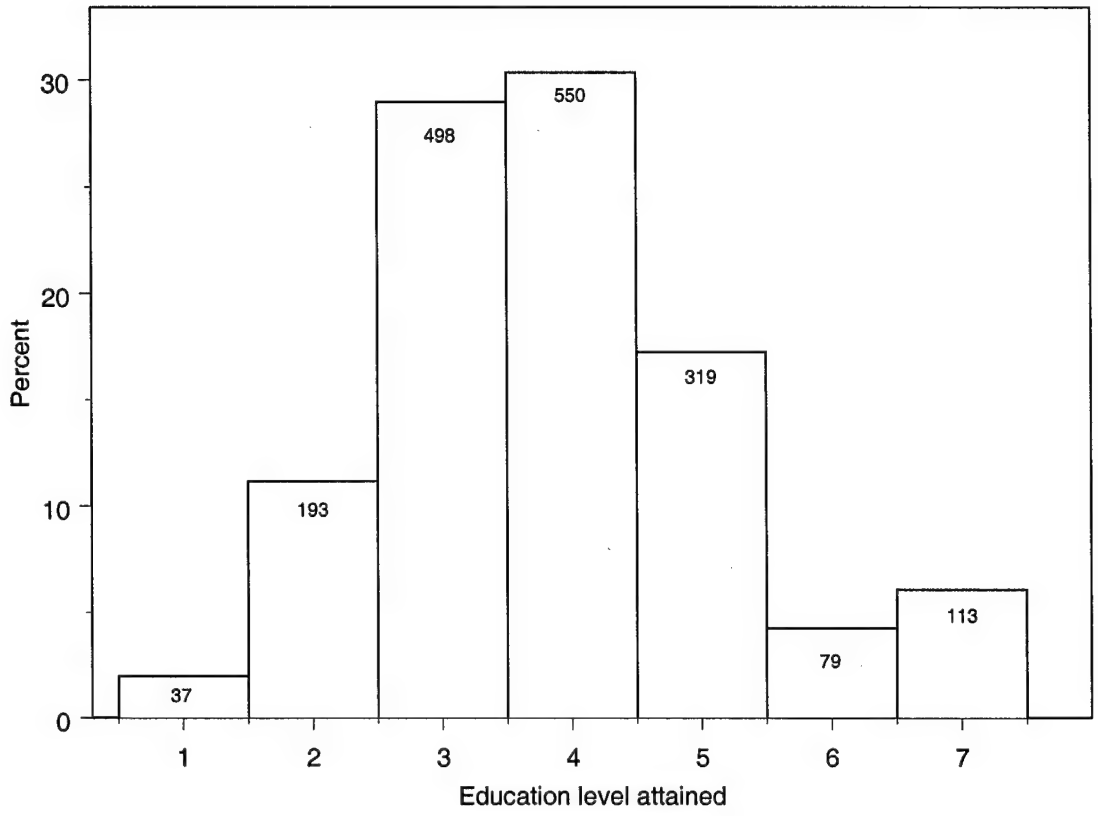
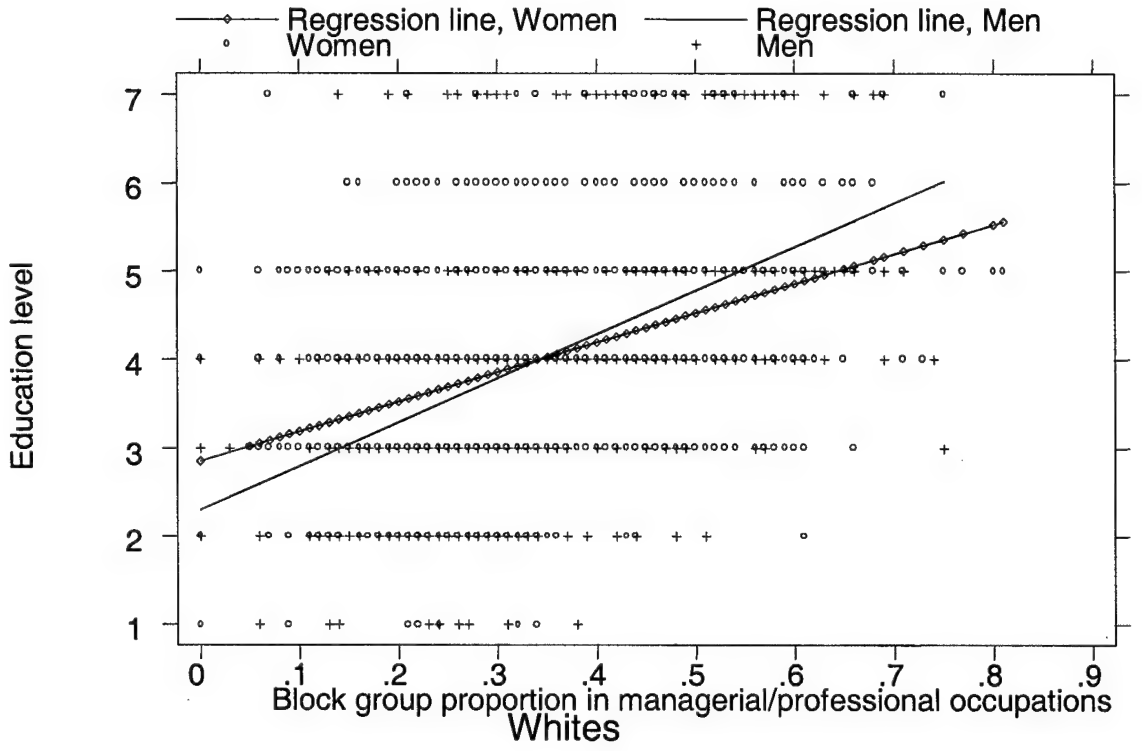


FIGURE 2. Regression lines describing the relationship between *individual education* and block group *proportion employed in managerial/professional occupations*, by gender and race



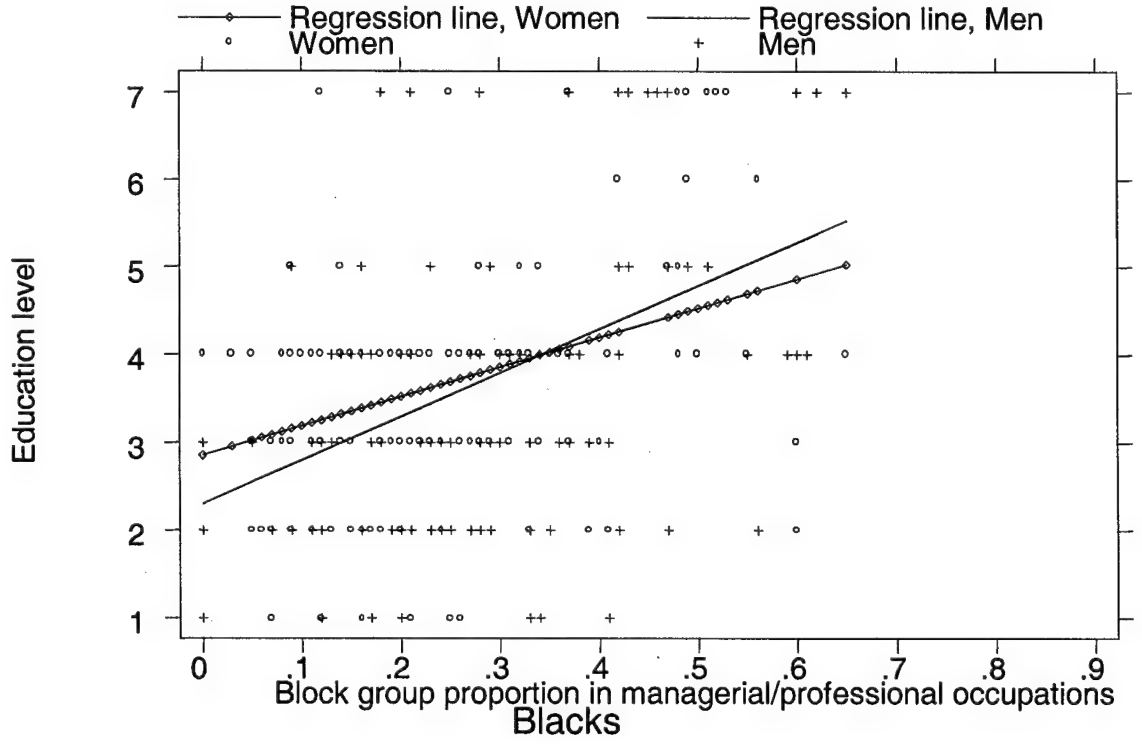
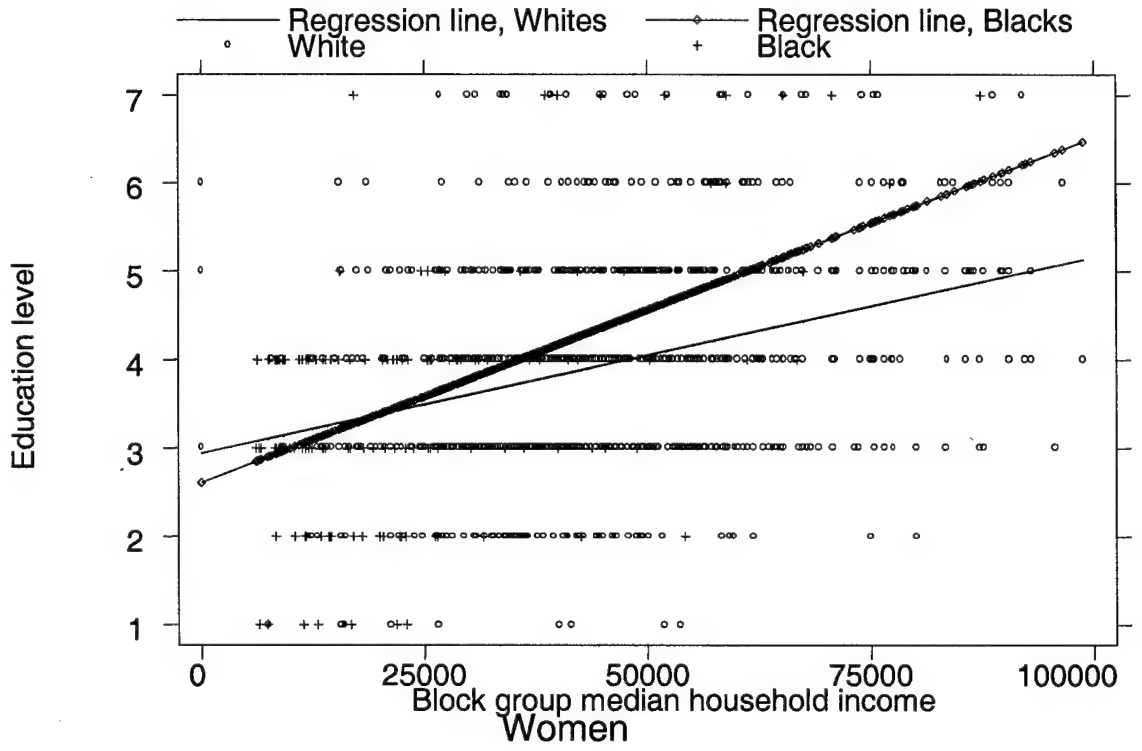


FIGURE 3. Regression lines describing the relationship between *individual education* and block group *median household income*, by race and gender



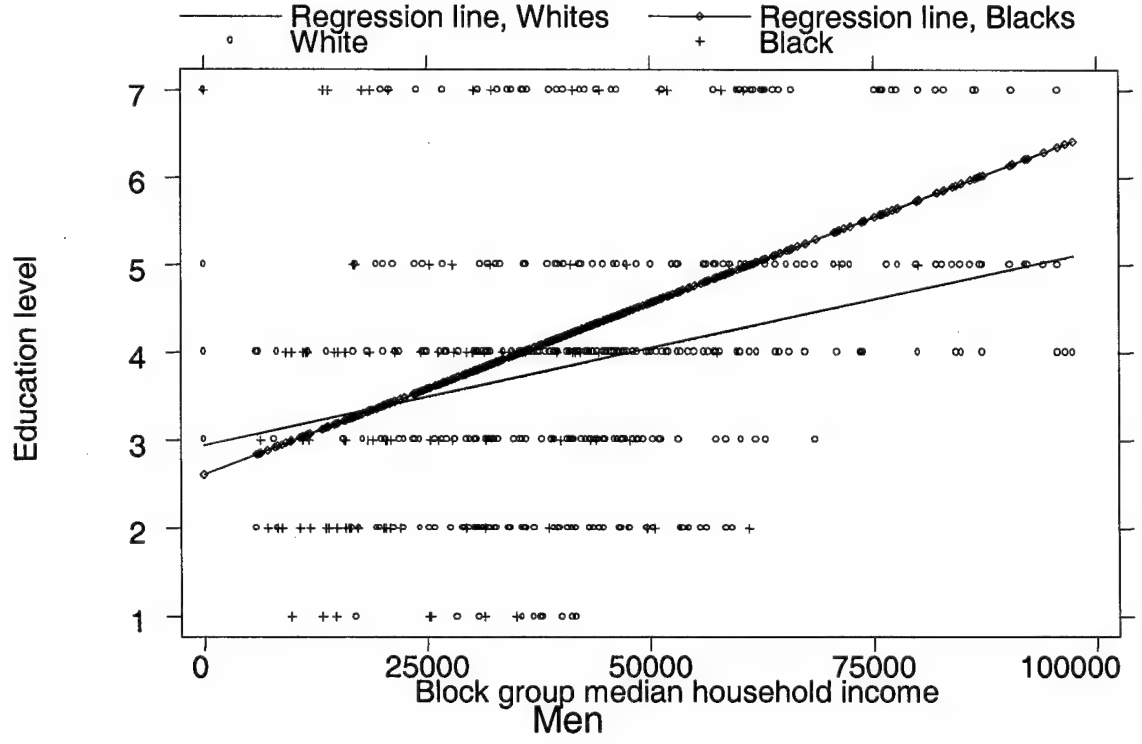


FIGURE 4. Loess spline regression estimate of mean education level by birth cohort

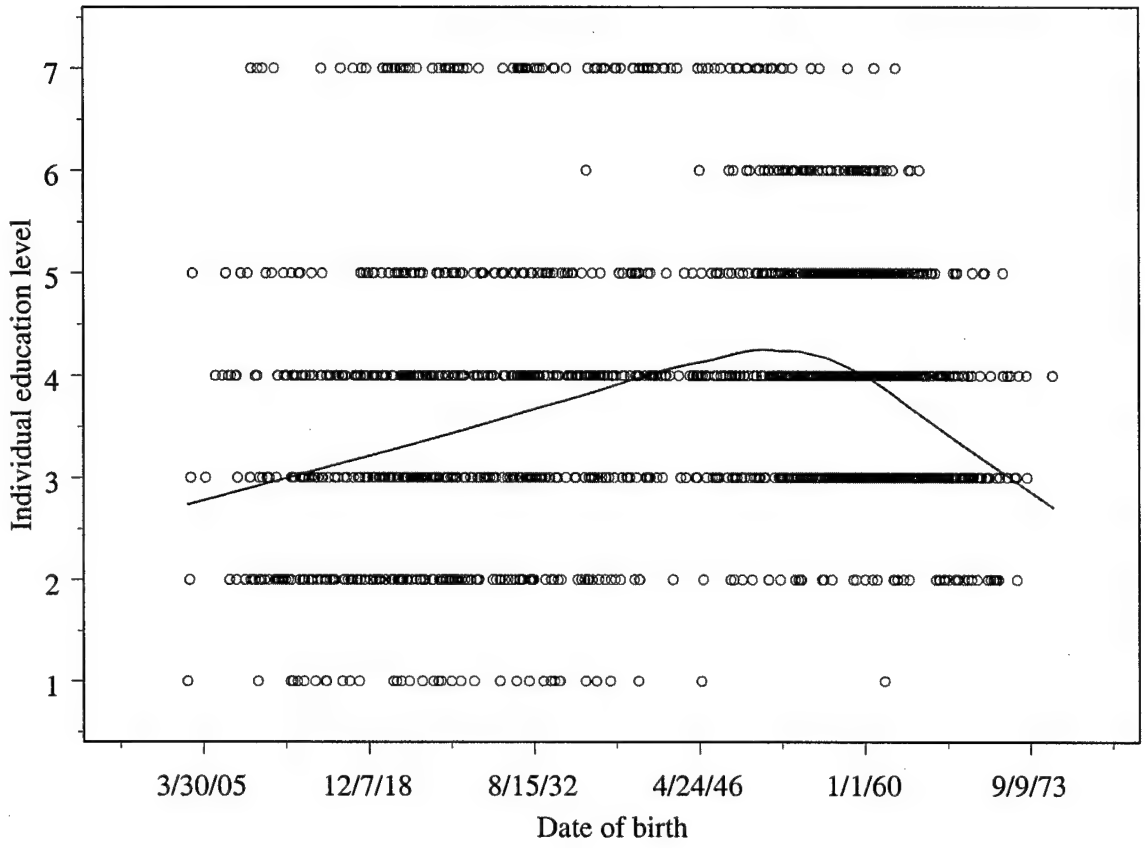
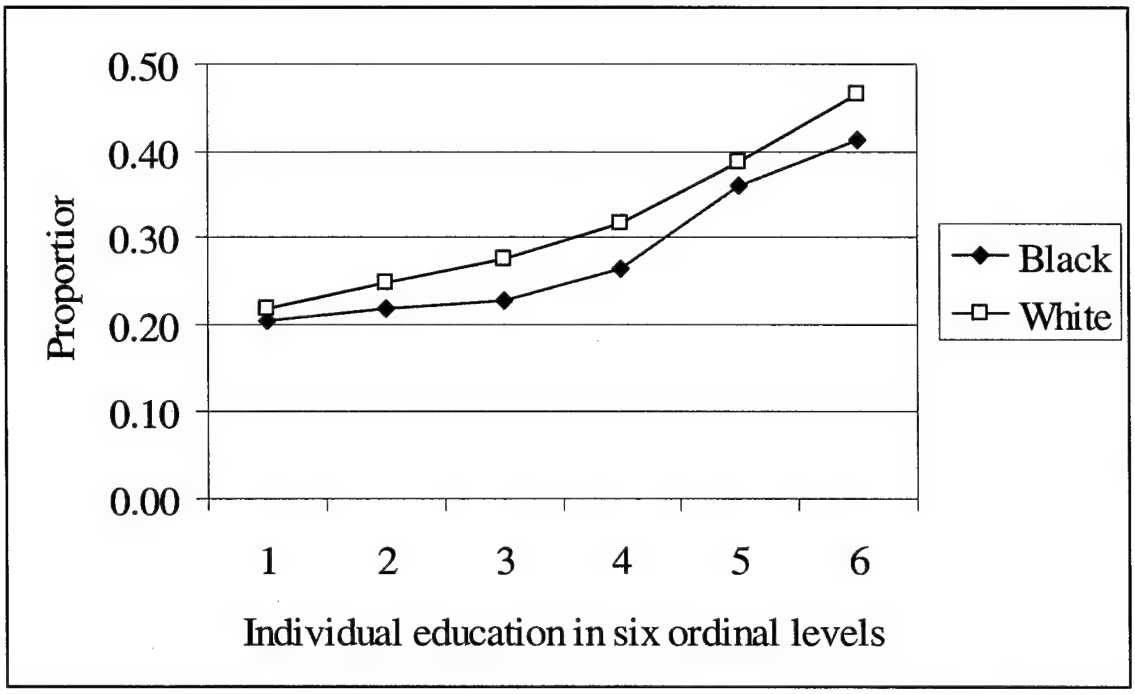
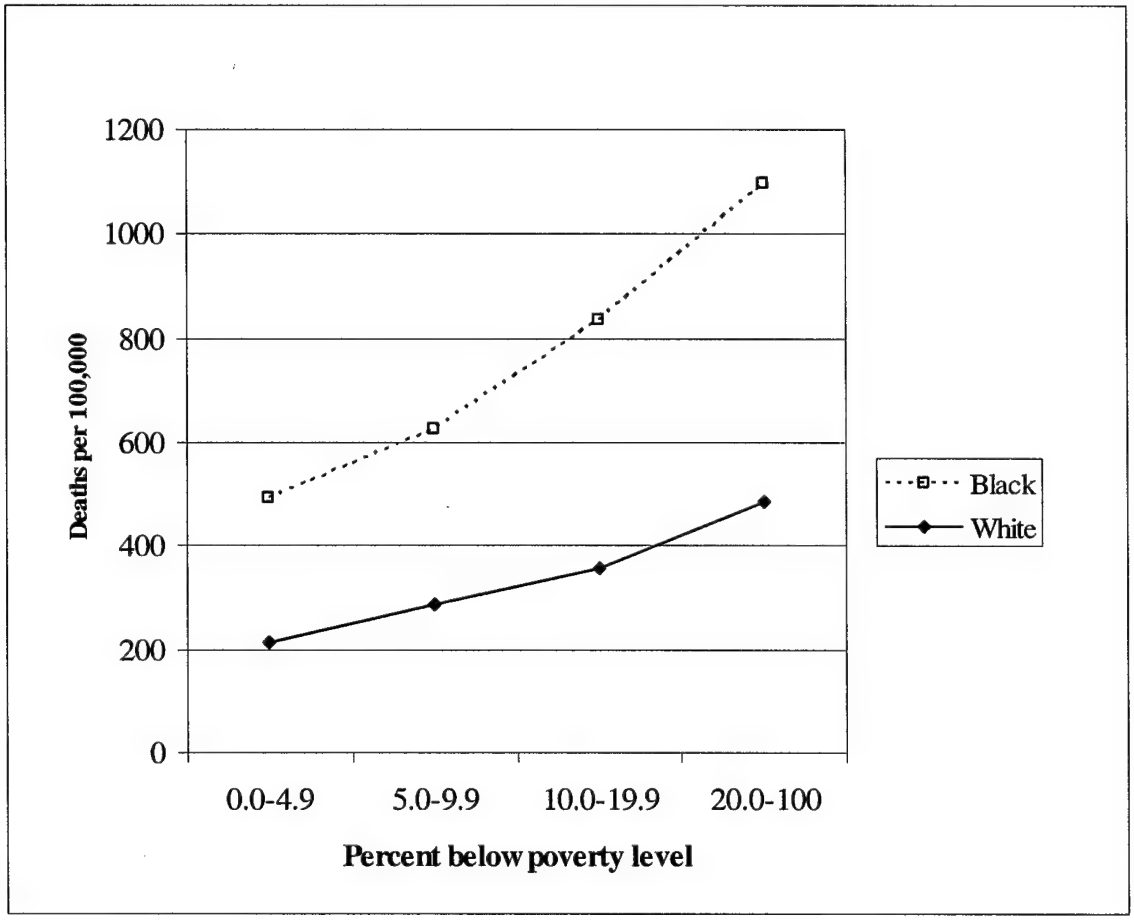


FIGURE 5. Mean block group *proportion in managerial/professional occupations* for six ordinal levels of individual education *, by race/ethnicity



* Because few blacks had incomplete post-graduate or professional education and this category was pooled together with college/university bachelor degree completed, yielding only 6 categories.

FIGURE 6. Premature male mortality (<65 years old) in Massachusetts, 1989-1991, by census tract percent below poverty, by race/ethnicity (Public Health Disparities Geocoding Project data plotted from data published by Krieger et al. 2003 (Krieger, Chen, Waterman et al., 2003))



APPENDIX 2

Comorbidity Classes

(1) INFECTIOUS AND PARASITIC DISEASES (ICD 001-139) No = 0 the default, YES = 1
(2) PREVIOUS NEOPLASMS (ICD 140-239)
(3) ENDOCRINE, NUTRITIONAL, METABOLIC & IMMUNITY DISORDERS (ICD 240-279)
(4) DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS (ICD 280-289)
(5) MENTAL DISORDERS (ICD 290-319)
(6) DISEASES OF THE NERVOUS SYSTEM & SENSE ORGANS (ICD 320-389)
- **CENTRAL NERVOUS SYSTEM**
- **PERIPHERAL NERVOUS SYSTEM**
- **SENSE ORGAN – EYE/OPHTHALMIC**
- **SENSE ORGAN – AUDITORY SYSTEM & OTHERS**
- **OTHER NERVOUS SYSTEM — Items not captured in preceding nervous system categories**

(7) DISEASES OF THE CIRCULATORY SYSTEM (ICD 390-459)
(8) DISEASES OF THE RESPIRATORY SYSTEM (ICD 460-519)
(9) DISEASES OF THE DIGESTIVE SYSTEM (ICD 520-579)
(10) DISEASES OF THE GENITOURINARY SYSTEM (580-629)
- **DISEASES OF URINARY TRACT**
- **DISEASES OF THE MALE GENITAL ORGANS**
- **DISEASES OF THE FEMALE GENITAL ORGANS**

(11) COMPLICATIONS OF PREGNANCY, CHILDBIRTH, & THE PUERPERIUM (IDC 630-677)
(12) DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE (ICD 680-709)
(13) DISEASES OF MUSCULOSKELETAL SYSTEM & CONNECTIVE TISSUE (ICD 710-739)
(14) CONGENITAL ANOMALIES (ICD 740-759)
(15) CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD (ICD 760-779)
(16) INJURY / TRAUMA & POISONING (800-999)

(17a) SYMPTOMS & SIGNS related to the index cancer, & ILL-DEFINED CONDITIONS (ICD 780-799)
(17b) SYMPTOMS & SIGNS unrelated to the index disease, & ILL-DEFINED CONDITIONS (ICD 780-799)

(1) INFECTIOUS AND PARASITIC DISEASES (ICD 001-139) No = 0 the default, YES = 1

CM1 **Tuberculosis.** Is this a recent infection (< 3 years old) or an active infection under treatment? ___
CM2 **Septicemia** (except in labor)
CM3 **Bacterial infection**, unspecified site
CM4 **Mycoses**
CM5 **HIV infection / AIDS**
CM6 **Hepatitis** (infectious, not primarily alcohol-related, see #150) *Circle*: Hepatitis virus A, B, C, D, E, G, or other.
CM7 **Viral infection** (not hepatitis)
CM8 **Other infections, including parasitic**
CM9 **Sexually transmitted infections** = STD (not HIV or hepatitis)
CM10 **(Immunizations and screening for infectious disease**, If yes, specify _____)
CM248 **Gangrene**

(2) PREVIOUS NEOPLASMS (ICD 140-239)

Cancer (CA) of	A. Present No=0, Yes=1	B. Metastasis No=0, Yes=1	C. Stage	D. Histology	E. Yr of diagnosis
CM11 Head & neck					
CM12 Esophagus					
CM13 Stomach					
CM14 Colon					
CM15 Rectum & anus					
CM16 Liver & intrahepatic bile duct					
CM17 Pancreas					
CM18 Other gastrointestinal organs, peritoneum					
CM19 Bronchus, lung					
CM20 Other respiratory & intra-thoracic					
CM21 Bone & connective tissue					
CM22 Melanomas of skin					
CM23 Other non-epithelial cancer of skin					
CM24 Breast					
CM25 Uterus					
CM26 Cervix					
CM27 Ovary					
CM28 Other female genital organs					
CM29 Prostate					
CM30 Testis					
CM31 Other male genital organs					
CM32 Bladder					
CM33 Kidney and renal pelvis					
CM34 Other urinary organs					
CM35 Brain and nervous system					
CM36 Thyroid					
CM37 Hodgkin's disease					
CM38 Non-Hodgkin's lymphoma					
CM39 Leukemias					
CM40 Multiple myeloma					
CM41 Other and unspecified primary					
CM42 Secondary malignancies					
CM43 Malignant neoplasm, unspecified site					
CM44 CA, unspecified/uncertain nature or behavior					
CM45 Maintenance chemotherapy, radiotherapy		N/A	N/A	N/A	N/A
CM46 Benign neoplasm of uterus, i.e., fibroids (leiomyoma; myoma; fibromyoma)		N/A	N/A		
CM47 Other and unspecified benign neoplasm		N/A	N/A		

(3) ENDOCRINE, NUTRITIONAL, METABOLIC & IMMUNITY DISORDERS (ICD 240-279)

CM48 Thyroid disorders e.g., goiter, hyperthyroidism, hypothyroidism, thyroiditis. If yes, specify ____
CM49 Diabetes mellitus without complication. If yes, is it insulin-dependent? Yes / No
CM50 Diabetes mellitus with complications. If yes, specify, e.g., ketoacidosis or uncontrolled diabetes, renal, ophthalmic, neurologic, circulatory, or other/unspecified complications. _____
If yes, is it insulin-dependent? Yes / No
CM51 Other endocrine disorders, e.g., parathyroid, pituitary and its hypothalamic control, adrenal or polyglandular disorders, premature ovarian failure (menopause <40years). If yes, specify _____
CM301 Obesity / hyperalimentation documented by physician/clinician/nurse in medical records
CM52 Nutritional deficiencies (specific). If yes, specify _____
CM52B Under-nutrition/malnutrition (general/unspecified)
CM53 Disorders of lipid metabolism, e.g., hypercholesterolemia, hyperlipidemia. If yes, specify ____
CM54 Gout and other crystal arthropathies, If yes, which of the following apply?
CM54B Gout, mild or not further specified
CM54C Gout with nephropathy
CM54D Gout with other specific manifestations
CM54E Other crystal arthropathy

CM55 Fluid and electrolyte metabolic disorders, If yes, please specify on table below (Circle and indicate Yes = 1)

Water balance	**CM55B Dehydration**	**CM55C Over-hydration**
Extracellular fluid volume	**CM55D Contraction**	**CM55E Expansion / Overload**
Sodium (Na)	**CM55F Hyponatremia**	**CM55G Hypernatremia**
Potassium (K)	**CM55H Hypokalemia (hypopotassemia)**	**CM55I Hyperkalemia (hyperpotassemia)**
Calcium (Ca)	**CM55J Hypocalcemia**	**CM55K Hypercalcemia**
Phosphate (P)	**CM55L Hypophosphatemia**	**CM55M Hyperphosphatemia**
Magnesium (Mg)	**CM55N Hypomagnesemia**	**CM55O Hypermagnesemia**
Acid-Base Metabolism	**CM55P Metabolic Acidosis** **CM55R Respiratory Acidosis**	**CM55Q Metabolic Alkalosis** **CM55S Respiratory Alkalosis**
Others, specify	**CM55T** _____	_____

CM302 Disorder of mineral metabolism, including iron, iodine, fluorine, zinc, chromium, selenium, manganese, molybdenum, & copper. If yes, specify _____
CM56 Cystic fibrosis
CM57 Immunity disorders, If yes, specify _____
CM253 Allergic reactions
CM303 Amyloidosis
CM58 Other nutritional, endocrine, and metabolic disorders, If yes, specify _____

(4) DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS (ICD 280-289)

CM59 Deficiency and other or unspecified anemia
CM60 Acute post-hemorrhagic anemia
CM61 Sickle cell anemia
CM62 Coagulation and hemorrhagic disorders
CM63 Diseases of white blood cells
CM64 Other hematologic conditions, including spleen disorders

(5) MENTAL DISORDERS (ICD 290-319)

CM65 Mental retardation
CM66 Alcohol-related mental disorders, including acute intoxication, dependency or abuse.
CM67 Substance-related mental disorders, including barbiturate, amphetamine, hallucinogen, opioid, cocaine or other or mixed drug dependence or abuse. Specify drugs _____
CM68 Senility & organic mental disorders, e.g., senile & arteriosclerotic dementia, Alzheimer's.
CM69 Affective disorders, e.g., depressive and bipolar affective disorder, manic-depressive psychosis.
CM70 Schizophrenia and related disorders
CM71 Other psychoses
CM72 Anxiety, somatoform, dissociative, and personality disorders
CM73 Pre-adult disorders
CM74 Other mental conditions
CM75 Personal history of mental disorder, mental & behavioral problems, observation/screening for mental condition

(6) DISEASES OF THE NERVOUS SYSTEM & SENSE ORGANS (ICD 320-389)

CENTRAL NERVOUS SYSTEM

CM76 Meningitis (except that caused by tuberculosis or sexually transmitted disease)
CM77 Encephalitis (except that caused by tuberculosis or sexually transmitted disease)
CM78 Other CNS infection and poliomyelitis If yes, specify _____
CM79 Parkinson's disease
CM80 Multiple sclerosis
CM81 Other hereditary & degenerative nervous system conditions, e.g., ALS. If yes, specify _____
CM82 Paralysis (except that secondary to cerebrovascular diseases which goes under # 113)
CM83 Epilepsy, convulsions
CM84 Headache, including migraine
CM85 Coma, stupor, and brain damage

PERIPHERAL NERVOUS SYSTEM

CM343 Peripheral neuropathy, unknown or specified etiology. If known, specify the cause? _____

SENSE ORGAN – EYE/OPHTHALMIC

CM86 Cataract
CM87 Retinal detachments, defects, vascular occlusion, and retinopathy
CM88 Glaucoma
CM89 Blindness and visually handicapped
CM90 Inflammation, infection of eye (except that caused by TB or STD)
CM337 Near-sightedness (myopia), far-sightedness (hyperopia), astigmatism or needing reading glasses (presbyopia)
CM91 Other eye/ ophthalmic disorders If yes, specify _____

SENSE ORGAN – AUDITORY SYSTEM & OTHERS

CM92 Otitis media and related conditions
CM93 Conditions associated with dizziness or vertigo
CM94 Other ear and sense organ disorders If yes, specify _____

OTHER NERVOUS SYSTEM — Items not captured in preceding nervous system categories

CM95 Other nervous system disorders If yes, specify _____

(7) DISEASES OF THE CIRCULATORY SYSTEM (ICD 390-459)

CM96 Heart valve disorders
CM97 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by tuberculosis or STD)
CM98 Essential hypertension
CM99 Hypertension with complications and secondary hypertension If yes, specify _____
CM100 Myocardial infarction How long ago was most recent MI? ____ years ____months prior to cancer diagnosis.
CM101 Coronary atherosclerosis and other heart disease
CM102 Angina (non-specific or non-angina chest pain is coded under #322)
CM103 Pulmonary heart disease (cor pulmonale)
CM340 Cardiomegaly
CM104 Other or ill-defined heart disease
CM105 Conduction disorders
CM106 Cardiac dysrhythmias / arrhythmias
CM107 Cardiac arrest or ventricular fibrillation
CM108 Congestive heart failure
CM109 Acute cerebrovascular disease
CM110 Occlusion or stenosis of precerebral arteries
CM111 Other and ill-defined cerebrovascular disease
CM112 Transient cerebral ischemia
CM113 Late effects of cerebrovascular disease, i.e., plegia or hemiplegia
CM114 Peripheral and visceral atherosclerosis
CM115 Aortic, peripheral, & visceral artery aneurysms,
- **CM115B** If yes, where was it located? _____.
- **CM115C** What was its size? _____ cm.
- **CM115D** Was it surgically corrected? No = 0, Yes = 1.

CM116 Aortic and peripheral arterial embolism or thrombosis
CM117 Other circulatory disease, including hypotension
CM118 Phlebitis, thrombophlebitis and thromboembolism
CM119 Varicose veins of lower extremity
CM120 Hemorrhoids
CM345 Lymphadenopathy
CM121 Other diseases of veins and lymphatics

(8) DISEASES OF THE RESPIRATORY SYSTEM (ICD 460-519)

CM122 Pneumonia (except that caused by tuberculosis or sexually transmitted disease)
CM123 Influenza
CM124 Acute and chronic tonsillitis
CM125 Acute bronchitis
CM126 Other upper respiratory infections, If yes, specify _____
CM127 Chronic obstructive pulmonary disease & bronchiectasis, If yes, specify:
- **CM127B COPD otherwise not specified**
- **CM127C Emphysema**
- **CM127D Chronic bronchitis**
- **CM127E Bronchiectasis or bronchiolectasis**

CM128 Asthma
CM304 Pulmonary fibrosis / interstitial lung diseases
CM129 Aspiration pneumonitis, food/vomitus
CM130 Pleurisy, pneumothorax, pulmonary collapse (atelectasis)
CM346 Pleural effusions, any cause
CM131 Respiratory failure, insufficiency, arrest (adult)
CM132 Lung disease due to external agents, including **pneumoconioses**, e.g., anthracosis, silicosis, asbestosis, berylliosis, siderosis, stannosis, & baritosis.
CM341 Sarcoidosis of the lung and including other non-pulmonary sites
CM133 Other lower respiratory disease
CM134 Other upper respiratory disease

(9) DISEASES OF THE DIGESTIVE SYSTEM (ICD 520-579)

CM135 Intestinal infection
CM136 Disorders of teeth and jaw
CM137 Diseases of mouth, excluding dental
CM138 Esophageal disorders
CM139 Gastroduodenal ulcer (except hemorrhage)
CM140 Gastritis and duodenitis
CM141 Other disorders of stomach and duodenum
CM142 Appendicitis and other appendiceal conditions
CM143 Abdominal hernia, If yes, was it accompanied by obstruction or gangrene? No = 0, Yes = 1.
CM144 Regional enteritis and ulcerative colitis, including inflammatory bowel diseases, such as Crohn's disease & ulcerative colitis.
CM145 Intestinal obstruction not from hernia, e.g., paralytic ileus, impaction, adhesions.
If yes, specify _____
CM342 Colorectal polyps, adenomatous polyps
CM146 Diverticulosis and diverticulitis
CM147 Anal and rectal conditions
CM148 Peritonitis and intestinal abscess
CM149 Biliary tract disease, e.g., cholecystitis, cholelithiasisis
CM150 Liver disease, alcohol-related
CM151 Other liver diseases, e.g., liver disease or cirrhosis without mention of alcohol, liver abscess.
CM152 Pancreatic disorders (not diabetes)
CM153 Gastrointestinal hemorrhage If yes, specify _____
CM154 Noninfectious gastroenteritis
CM155 Other gastrointestinal disorders, e.g., constipation, dysphagia. If yes, specify _____

(10) DISEASES OF THE GENITOURINARY SYSTEM (580-629)

CM156 Nephritis, nephrosis, renal sclerosis, If yes, specify _____
CM157 Acute and unspecified renal failure
CM158 Chronic renal failure
CM335 Has the patient had dialysis? If yes, earliest date _____ and last date _____
CM159 Urinary tract infections, If yes, specify if of kidney or cystitis/urethritis : _____
CM160 Calculus of urinary tract (urolithiasis) If yes, specify if of kidney or ureter or bladder : _____
What is the composition?: calcium oxalate; uric acid; cystine; struvite = magnesium ammonium phosphate, other, unknown.
CM161 Other diseases of kidney and ureters, e.g., hydronephrosis
CM162 Other diseases of bladder and urethra
CM163 Genitourinary symptoms & ill-defined conditions, e.g., hematuria, dysuria, retention of urine.

DISEASES OF THE MALE GENITAL ORGANS
CM164 Hyperplasia of prostate
CM165 Inflammatory conditions of male genital organs, If yes, specify _____
CM166 Other male genital disorders, If yes, specify _____

DISEASES OF THE FEMALE GENITAL ORGANS
CM167 Nonmalignant breast conditions
CM168 Inflammatory diseases of female pelvic organs, e.g., pelvic peritoneal adhesions, cervicitis / endocervicitis, pelvic inflammatory disease (including endometritis, salpingitis and ooporitis). Specify ___
CM169 Endometriosis
CM170 Prolapse of female genital organs
CM171 Menstrual disorders
CM172 Ovarian cyst
CM173 Menopausal disorders
CM174 Female infertility
CM175 Other female genital disorders

(11) COMPLICATIONS OF PREGNANCY, CHILDBIRTH, & THE PUERPERIUM (IDC 630-677)

CM176 Contraceptive and procreative management
CM177 Spontaneous abortion
CM178 Induced abortion
CM179 Post-abortion complications
CM180 Ectopic pregnancy
CM181 Other complications of pregnancy, e.g., genitourinary infection during pregnancy, anemia during pregnancy, mental disorder during pregnancy, missed abortion, hyperemesis gravidarum, infectious/parasitic complications in mother affecting pregnancy. If yes, specify _____
CM182 Hemorrhage during pregnancy, abruptio placenta, placenta previa
CM183 Hypertension complicating pregnancy, childbirth and the puerperium, e.g., preeclampsian/eclampsia.
CM184 Early or threatened labor
CM185 Prolonged pregnancy
CM186 Diabetes or abnormal glucose tolerance complicating pregnancy, childbirth, or the puerperium
CM187 Malposition, malpresentation
CM188 Fetopelvic disproportion, obstruction
CM189 Previous cesarean section
CM190 Fetal distress and abnormal forces of labor, e.g., fetal distress, uterine inertia, precipitate labor.
CM191 Polyhydramnios & other problems of amniotic cavity. e.g., premature rupture of membranes, infection of amnion.
CM192 Umbilical cord complication
CM193 Trauma to perineum and vulva
CM194 Forceps delivery
CM195 Other complications of birth, puerperium affecting management of mother, e.g., postpartum hemorrhage, cervical incompetence, rhesus isoimmunization, interuterine death, failed induction.
CM196 Normal pregnancy and/or delivery

(12) DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE (ICD 680-709)

(Include in this category diseases of structures developed from skin, such as toe and finger nails.)

CM197 Skin and subcutaneous tissue infections, e.g., cellulitis or abscess.
CM198 Other inflammatory condition of skin
CM199 Chronic ulcer of skin
CM200 Other skin disorders

(13) DISEASES OF MUSCULOSKELETAL SYSTEM & CONNECTIVE TISSUE (ICD 710-739)

CM201 Infective arthritis and osteomyelitis (except that caused by TB or STD)
CM202 Rheumatoid arthritis and related disease
CM203 Osteoarthritis
CM204 Other non-traumatic joint disorders (place gout and other crystalline metabolic arthropathic disorders in #54)
CM205 Spondylosis, intervertebral disc disorders, other back problems
CM206 Osteoporosis
CM206B Osteopenia
CM207 Pathological fracture
CM208 Acquired foot deformities
CM209 Other acquired deformities
CM210 Systemic lupus erythematosus and connective tissue disorders
CM211 Other connective tissue disease
CM212 Other bone disease and musculoskeletal deformities
CM305 Limb amputation, If yes, then check if #254 applies.
CM339 Hip replacement

(14) CONGENITAL ANOMALIES (ICD 740-759)

CM213 Cardiac and circulatory congenital anomalies
CM214 Digestive congenital anomalies
CM215 Genitourinary congenital anomalies
CM216 Nervous system congenital anomalies
CM217 Other congenital anomalies

(15) CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD (ICD 760-779)

CM218 Liveborn
CM219 Short gestation, low birth weight, and fetal growth retardation
CM220 Intrauterine hypoxia and birth asphyxia
CM221 Respiratory distress syndrome
CM222 Hemolytic jaundice and perinatal jaundice
CM223 Birth trauma
CM224 Other perinatal conditions

(16) INJURY / TRAUMA & POISONING (800-999)

CM225 Joint disorders and dislocations, trauma-related
CM226 Fracture of neck of femur (hip)
CM227 Spinal cord injury
CM228 Skull and face fractures
CM229 Fracture of upper limb
CM230 Fracture of lower limb
CM231 Other fractures
CM232 Sprains and strains
CM233 Intracranial injury
CM234 Crushing injury or internal injury
CM235 Open wounds of head, neck, and trunk
CM236 Open wounds of extremities
CM237 Complication of device, implant or graft
CM238 Complications of surgical procedures or medical care
CM239 Superficial injury, contusion
CM240 Burns
CM241 Poisoning by psychotropic agents
CM242 Poisoning by other medications and drugs
CM243 Poisoning by nonmedicinal substances
CM244 Other injuries and conditions due to external causes
CM306 Gunshot injury

(17a) SYMPTOMS & SIGNS related to the index cancer, & ILL-DEFINED CONDITIONS (ICD 780-799)

CM307A Did the patient present with symptoms related to the cancer under study? No=0, Yes=1.
CM307B If yes, what was the duration of symptoms? days __ weeks __ months__ years__.
If symptomatic, complete the table below.

GENERAL	**CM245 Syncope, fainting** **CM249 Shock** **CM252 Fatigue and malaise, i.e., tiredness, weakness, lethargy** **CM246 Fever**, tumor-related or of unknown origin **CM308 Chills, sweats, night sweats, diaphoresis (excess or profuse perspiration)** **CM309 Weight loss (unintentional)** How many pounds were lost? ___, Over how many months? ____. Was weight loss intentional (i.e., due to dieting)? = 0, or was it disease related? = 1
GASTRO-INTESTINAL	**CM250 Nausea, vomiting, emesis** **CM310 Anorexia, loss of appetite, decreased appetite** **CM311 Heartburn** **CM336 Jaundice, icterus**
RESPIRA-TORY / CHEST	**CM312 Upper respiratory symptoms, epistaxis** **CM313 Throat symptoms, e.g.,** dysphagia, difficulty swallowing, sore throat, swollen throat, hiccups, choking sensation, hoarseness (rough or harsh quality of voice), dysphonia (any impairment of voice, a difficulty in speaking) **CM314 Cough** **CM315 Dyspnea, shortness of breath (SOB), excertional dyspnea, orthopnea** (inability to breath except in an upright position) **CM316 Wheezing** (i.e., whistling noises, high pitch, made during breathing) or **Stridor** (a harsh sound, audible without a stethoscope and predominantly inspiratory, often from obstruction) **CM317 Respiratory congestion** **CM318 Palpitations** **CM319 Hemoptysis** (coughing up blood from the respiratory tract) **CM320 Cyanosis** **CM321 Finger clubbing**
PAIN	**CM251 Abdominal pain** **CM322 Chest pain other than angina** **CM323 Pain of the back** **CM324 Pain of the shoulder** **CM325 Other pain**, e.g., arthralgia, neuralgia, pain in extremities.
NODES, MASSES, SWELLINGS	**CM247 Lymphadenitis** **CM326 Lymphadenopathy or palpable mass** or "can feel mass". **CM327 Swelling / edema**
NEURO-MUSCULAR & MENTAL	**CM328 Headache** as a presenting sign/symptom of the index cancer **CM329 Diziness** **CM330 Eye / ophthalmic symptoms & signs**, e.g., blurred vision, diplopia, photophobia. **CM331 Dysmetria** (improper measuring of distance or range of movement in muscular action) **CM338 Insomnia** **CM332 Mental changes** as a presenting sign/symptom of the index cancer **CM333 Neurologic symptoms & signs** as a presenting sign/symptom of the index cancer
OTHER	**CM334 Alopecia, hair loss** **CM344a Speech defect, disorder, disturbance, impediment**. Is this a recent change (last 1 years)? ___ **CCM347 Polydipsia** **CCM348 Polyurea**

CM254 Rehabilitation care, fitting of prostheses, and adjustment of devices

(17b) SYMPTOMS & SIGNS unrelated to the index disease, & ILL-DEFINED CONDITIONS (ICD 780-799)

CM307Ab Did the patient have symptoms **unrelated** to the cancer under study? No=0, Yes=1.
CM307Bb If yes, how long ago did they start? _____ For how long did they last _____?

If a history of symptoms occurred in the five years prior to diagnosis with the index cancer, complete the table below.

GENERAL	**CM245b Syncope, fainting** **CM249b Shock** **CM252b Fatigue and malaise, i.e., tiredness, weakness, lethargy** **CM246b Fever**, tumor-related or of unknown origin **CM308b Chills, sweats, night sweats, diaphoresis (excess or profuse perspiration)** **CM309b Weight loss (unintentional)** How many pounds were lost? ___, Over how many months? ____. Was weight loss intentional (i.e., due to dieting)? = 0, or was it disease related? = 1
GASTRO-INTESTINAL	**CM250 Nausea, vomiting, emesis** **CM310 Anorexia, loss of appetite, decreased appetite** **CM311 Heartburn** **CM336 Jaundice, icterus**
RESPIRA-TORY / CHEST	**CM312 Upper respiratory symptoms, epistaxis** **CM313 Throat symptoms, e.g.,** dysphagia, difficulty swallowing, sore throat, swollen throat, hiccups, choking sensation, hoarseness (rough or harsh quality of voice), dysphonia (any impairment of voice, a difficulty in speaking) **CM314 Cough** **CM315 Dyspnea, shortness of breath (SOB), excertional dyspnea, orthopnea** (inability to breath except in an upright position) **CM316 Wheezing** (i.e., whistling noises, high pitch, made during breathing) or **Stridor** (a harsh sound, audible without a stethoscope and predominantly inspiratory, often from obstruction) **CM317 Respiratory congestion** **CM318 Palpitations** **CM319 Hemoptysis** (coughing up blood from the respiratory tract) **CM320 Cyanosis** **CM321 Finger clubbing**
PAIN	**CM251 Abdominal pain** **CM322 Chest pain other than angina** **CM323 Pain of the back** **CM324 Pain of the shoulder** **CM325 Other pain**, e.g., arthralgia, neuralgia, pain in extremities.
NODES, MASSES, SWELLINGS	**CM247 Lymphadenitis** **CM326 Lymphadenopathy or palpable mass** or "can feel mass". **CM327 Swelling / edema**
NEURO-MUSCULAR & MENTAL	**CM328 Headache** as a presenting sign/symptom of the index cancer **CM329 Diziness** **CM330 Eye / ophthalmic symptoms & signs**, e.g., blurred vision, diplopia, photophobia. **CM331 Dysmetria** (improper measuring of distance or range of movement in muscular action) **CM338 Insomnia** **CM332 Mental changes** as a presenting sign/symptom of the index cancer **CM333 Neurologic symptoms & signs** as a presenting sign/symptom of the index cancer
OTHER	**CM334 Alopecia, hair loss** **CM344b Speech defect, disorder, disturbance, impediment.** Is this a recent change (last 10 years)? ___ **CCM347 Polydipsia** **CCM348 Polyurea**

(17) UNCLASSIFIED, continued

CM259 Residual codes, unclassified

Other: Describe _____

ADDITIONS

CM337 Myopia, hyperopia, astigmatism, presbyopia, added Jan 30, 2002
CM338 Insomnia, added Jan 30, 2002
CM339 Hip replacement, added Jan 30, 2002
CM127E Bronchiectasis or bronchiolectasis, added November 20, 2002.
CM340 Cardiomegaly, added December 11, 2003.
CM341 Sarcoidosis of lung plus other non-pulmonary sties, added December 11, 2003.
CM342 Colorectal polyps, adenomatous polyps, added December 11, 2003.
CM343 Peripheral neuropathy, unknown or specified etiology, added December 11, 2003.
CM344a & b. Speech defect, disorder, disturbance, impediment. Is this a recent change (last 1 years)? added December 11, 2003.
CM345 Lymphadenopathy, added December 11, 2003.
CM346 Pleural effusions, any cause, added December 11, 2003.
CCM347 Polydipsia, added December 11, 2003.
CCM348 Polyurea, added December 11, 2003.